Variational Structures for Dynamical Fluctuations, In and Out of Equilibrium (Alternative Format Thesis)

Kaiser, Marcus

Award date:
2018

Awarding institution:
University of Bath

Variational Structures for Dynamical Fluctuations, In and Out of Equilibrium

submitted by

Marcus Kaiser

for the degree of *Doctor of Philosophy*

of the

University of Bath

Department of Mathematical Sciences

July 2018

Declaration of any previous Submission of the Work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

. . . . . .

Marcus Kaiser

Declaration of Authorship

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of Chapters 2 to 4, which contain research articles that originated from collaboration with my supervisors Robert L. Jack and Johannes Zimmer.

. . . . . .

Marcus Kaiser

Summary

In this thesis, we investigate variational structures for fluctuations in Markov processes, with a particular focus on interacting particle systems (such as the simple exclusion process and the zero-range process).

A great part of this thesis is devoted to time-reversal symmetry. We discuss the acceleration of convergence to the steady state for dissipative systems, where we revisit the fact that 'breaking detailed balance' accelerates the convergence to equilibrium and extend known results to the case of interacting particle systems and their hydrodynamic scaling limits. The theoretical findings are supported by simulations of independent particles and the zero-range process in one and two space dimensions.

We further investigate a general Ψ-$\Psi^\star$ structure for the Onsager-Machlup functional Φ, which can be used to represent several large-deviation rate functions for particle diffusions, Markov chains and Macroscopic Fluctuation Theory. We discuss a splitting of the thermodynamic force acting on the system in time-reversal symmetric and anti-symmetric parts, for which we prove a 'generalised Hamilton-Jacobi orthogonality'.

Finally, we apply this structure to a special class of interacting particle systems (which includes the simple-exclusion process and a large class of zero-range processes) and show how the individual terms of the Ψ-$\Psi^\star$ structure converge to their hydrodynamic counterparts (as known from Macroscopic Fluctuation Theory).

Acknowledgements

First and foremost, I want to thank my supervisors Johannes and Rob for their guidance and continued support of my PhD study. I very much enjoyed all the discussions we had over the last three years.

Further, I would like to thank all of my friends at the Department of Mathematical Sciences at the University of Bath and the SAMBa team. In particular, I am indebted to Federico Cornalba and Marios Stamatakis for many mathematical discussions. I also want to thank my examiners Alexander M.G. Cox and Mark A. Peletier for carefully reading the thesis and the helpful suggestions they made for the final version.

This research was supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/L015684/1.

Contents

1 Introduction **1**
- 1.1 Microscopic, Mesoscopic and Macroscopic Processes . . . 2
- 1.2 Equilibrium vs. Non-Equilibrium Systems . . . 3
  - 1.2.1 Classical Thermodynamics . . . 3
  - 1.2.2 Time-Reversal Symmetry . . . 3
  - 1.2.3 Persistent currents in the steady state . . . 3
  - 1.2.4 Time-Reversal Symmetric Processes as Gradient Flows . . . 4
- 1.3 Large Deviation Theory . . . 5
  - 1.3.1 LDPs for Markov Chains . . . 6
  - 1.3.2 LDPs on the Macroscopic Scale . . . 6
- 1.4 Outline of the Thesis . . . 7

2 Acceleration of Convergence to Equilibrium in Markov Chains by Breaking Detailed Balance **9**
- 2.1 Outline of the Article . . . 9
- 2.2 Conclusions . . . 40

3 Canonical Structure and Orthogonality of Forces and Currents in Irreversible Markov Chains **41**
- 3.1 Outline of the Article . . . 41
- 3.2 Conclusions . . . 75

4 A Variational Structure for Interacting Particle Systems and their Hydrodynamic Scaling Limits **77**
- 4.1 Outline of the Article . . . 77
- 4.2 Conclusions . . . 118

5 Final Conclusions and Outlook **119**
- 5.1 Outlook . . . 119

Chapter 1

Introduction

Non-equilibrium phenomena are ubiquitous and deeply linked to the physical behaviour of the real world. Systems out of equilibrium describe a vast number of processes on various length scales. (Precise definitions of these quantities will be given in Section 1.2 below.) Examples reach e.g. from the conversion of chemical energy into mechanical motion in molecular motors (on the nano-scale) [33, 26], to the occurrence of extreme weather events in meteorology (on the synoptic scale) [45, 27].

Jona-Lasinio [27] noted that, despite the dominant role of non-equilibrium processes, the current understanding of these processes is often based on concepts, originally derived for physical systems in (or very close to) equilibrium, which are not valid *far-from-equilibrium*.

A striking success in the theory of non-equilibrium systems are *Fluctuation Theorems (FTs)* [8], such as the *Gallavotti-Cohen symmetry* [19], *Jarzynski's identity* [25] and *Crooks Fluctuation Theorem* [9]. This is a class of closely related results that can, for example, relate the probability of a sample path to its time-reversed trajectory via the rate of *entropy production* along the trajectory, or yield a non-equilibrium work relation, which relates the work performed in the system to the change in free energy [26]. Note that the latter links a non-equilibrium quantity (the work performed on the system) to an equilibrium quantity (the free energy). We refer the reader to [15] for a detailed review of several FTs. One reason for the importance of these results lies in the fact that the established relations hold arbitrarily far-from-equilibrium. The *Differential Fluctuation Theorem* (which implies several other FTs, such as Jarzynski's identity and Crooks Fluctuation Theorem) has recently been confirmed experimentally [21] for the first time. For example, FTs were used to analyse the *internal thermal noise* in test masses for the LIGO observatory [36], which was recently used to discover *gravitational waves*.

Another seminal result, which will play a dominant role in this thesis, is *Macroscopic Fluctuation Theory (MFT)*, which is a unified theory for the macroscopic description of diffusive systems out of equilibrium [27, 2]. As we will see later, this theory allows to derive thermodynamic quantities, such as the *quasipotential* (which can be interpreted as a *non-equilibrium free energy*).

We argue that a better understanding of non-equilibrium systems is needed in order to gain new insights into areas such as *active matter* [7], which can e.g. model 'living systems', such as bird flocks or bacteria. These systems consist of a large number of self-driven constituents (in the above examples birds or bacteria) that dissipate energy, such that these systems are far from equilibrium. This implies that general principles from equilibrium thermodynamics, such as the *Fluctuation Dissipation Theorem (FDT)*, are no longer valid, see e.g. [17]. In active matter models, such as self-propelled particles (e.g. *Run-and-Tumble Particles* or *Active Brownian Particles*), one can observe *collective behaviour*, such as flocking and *Motility-Induced Phase Separation (MIPS)*, which cannot be observed for simpler models, such as Brownian particles (see again [7]). To the present day, there is no general theory for active matter, such that one often has to rely on computational results, as only few models can be solved exactly, see e.g. [34].

1.1. Microscopic, Mesoscopic and Macroscopic Processes

For our later considerations, it is useful to introduce three different length scales. The smallest length scale (above the quantum realm, which we will not consider in this thesis) is the *microscopic scale*, on which one is mainly concerned with the evolution of individual particles, such as atoms or molecules, described in terms of the laws of *classical mechanics (Newton's laws of motion)*. An example is given by *molecular dynamics*, where the (deterministic) evolution of an isolated system is usually described by *Hamiltonian mechanics*. If the system is not completely isolated, there is an exchange of some kind of energy with the environment. One way of modelling this coupling is using thermostats, which introduces randomness into the system, which can for example be modelled using *Langevin dynamics*, see e.g. Section 2.3 in [26]. Thus microscopic systems are models with a high level of detail, which can be both deterministic or stochastic.

At the other end of the spectrum we find the *macroscopic scale*, on which one deals with the evolution of *macroscopic quantities*, such as the flow of heat or a density of particles. This can be seen as a simplified, or *coarse-grained*, description of the underlying microscopic system, where one neglects the dynamics of individual particles in favour of macroscopic densities that evolve according to appropriate conservation laws from e.g. *continuum mechanics*. In this work, we will focus on diffusive systems, as described by *Macroscopic Fluctuation Theory (MFT)* [2], for which the evolution of a particle density ρ is given by a conservation law of the form $\dot{\rho}_t = -\nabla \cdot J(\rho_t)$ for a current $J(\rho) = -D(\rho)\nabla\rho + \chi(\rho)E$ (leading to a non-linear diffusion equation).

Finally, the *mesoscopic scale* can be thought of as an intermediate scale that interpolates between the microscopic and the macroscopic scale: It is a coarse-grained description of the microscopic dynamics, which at the same time is more detailed than the situation on the macroscopic scale. These systems are usually random and thus can conveniently be described using methods from *statistical mechanics* [8].

Our main focus will lie on *lattice gasses*, a special class of *interacting particle systems*. These systems are described by continuous time (finite or countable state) Markov chains and thus can be seen as mesoscopic systems. Prominent examples are the *Simple Exclusion Process (SEP)* and the *Zero-Range Process (ZRP)* [32]. We will further be interested in the associated *hydrodynamic scaling limits*, which yield macroscopic descriptions in terms of MFT.

For scaling limits, one usually considers two scales, a "more microscopic" and a "more macroscopic" scale [20], where one is generally interested in coarse graining the more microscopic scale, which leads to a description of the system on the more macroscopic scale. Since we are only interested in two scales, it is common to refer to particle systems, despite being defined on the mesoscopic scale, as microscopic systems and to the coarse grained systems as macroscopic systems. We will adapt this notation. This should not lead to any confusion, since we will not consider microscopic systems (in the sense as defined above).

In the following, we will mostly be concerned with *dissipative processes*, i.e. processes that relax to a steady state in the long time limit (as $t \to \infty$). This steady state will assumed to be unique and thus globally attractive.

Note that results similar to the ones considered in this thesis, which are related to dissipation in non-equilibrium systems, have recently been obtained in [23] for (both over- and underdamped) Langevin dynamics.

1.2. Equilibrium vs. Non-Equilibrium Systems

1.2.1 Classical Thermodynamics

Let us describe what we mean with equilibrium and non-equilibrium systems. The best starting point is classical *equilibrium thermodynamics* (see e.g. Chapter 1 in [8]). A macroscopic system is said to be in *thermodynamic equilibrium* if the system relaxed (after waiting for a possibly very long time) to a stationary state, with no net flow of macroscopic quantities (neither within the system, nor in exchange with the environment). In other words, the *thermodynamic variables* describing the system, such as *internal energy, temperature, volume,* and *number of particles,* are in thermodynamic equilibrium assumed to be conserved and uniform over the system; for example there is a unique constant temperature characterising the whole system.

From the thermodynamical perspective, any system that is not in thermodynamic equilibrium represents a non-equilibrium system. A very important special case is given by *non-equilibrium steady states*. This is the case where the system eventually relaxes to a stationary state, such that the values of the macroscopic variables do not change as time evolves, but there is a balanced exchange (i.e. no net flow) of e.g. heat between the system and its environment. A simple example is given by an iron rod that is coupled on both ends to different heat baths at different temperatures (cf. [27]), leading to a steady heat flow through the rod. Clearly, in this case the temperature in the system is no longer uniform, which leads to the key concept of *local equilibrium*. Loosely speaking, a macroscopic non-equilibrium system is said to be in local equilibrium if the system consists of many small subsystems, each of them in (or close to) equilibrium, with locally well defined thermodynamic quantities. We can think of a system in local equilibrium as a system that can locally, at any point, be described in terms of macroscopic densities, such that for example the temperature of the system is locally well-defined. Note that this assumption also lies at the heart of MFT, which allows us to define a particle density ρ in first place. Without a local equilibrium assumption, quantities such as temperatures and (particle) densities may, in general, be ill-defined.

1.2.2 Time-Reversal Symmetry

The type of processes we want to consider here (dissipative processes with unique steady states) are naturally characterised by their steady states. We consider two types of processes: time-reversal symmetric (or reversible, for short) and irreversible ones. *Reversible processes* have an equilibrium steady state (with no net flow in the variables that describe the process), whereas *irreversible processes* have non-equilibrium steady states (which exhibit persistent currents).

Let us first consider the macroscopic case, i.e. MFT. Denoting the steady state with $\bar{\rho}$, a system is time-reversal symmetric if and only if $J(\bar{\rho}) = 0$ for the current introduced above.

For Markov chains, we can describe the state of the process by the *probability distribution*. Equilibrium processes are then characterised by a vanishing *probability current* (which will be introduced below). We stress that this choice is in accordance with the general theory of *stochastic thermodynamics* (see [49] and also [48]).

1.2.3 Persistent currents in the steady state

We consider a basic example which explains the general idea of probability currents in stochastic models and the emergence of persistent currents in non-equilibrium steady

states. One of the simplest examples one can think of is a single particle travelling on a circle: We consider a simple random walk on a flat torus with L sites (in continuous-time and with homogeneous nearest neighbour transition rates). The configuration space (also called state space) can then be defined to be $\Omega = \{0, \dots, L-1\}$. (Alternatively, one could consider a single particle diffusion on $[0, 2\pi]$ where the boundary points are identified.) We denote the rate for the particle to jump from site i to site $i+1$ with r_+ and the rate to jump from $i+1$ to i with r_-. For any choice of the rates r_+ and r_-, this process is ergodic w.r.t. the flat (or uniform) distribution $\pi(i) = 1/L$: after a very long time, the particle has forgotten about its initial distribution and it is equally likely to be found at any of the L sites.

When $r_+ > r_- \geq 0$, the particle has a, say, counter-clockwise drift that forces the particle to travel around the circle. This choice of rates corresponds to an irreversible process, whereas the symmetric simple random walk (with $r_+ = r_- > 0$) is reversible.

For Markov chains, time-reversal symmetric processes are classically characterised by the *detailed balance* equation $\pi(i)r_+ = \pi(i+1)r_-$. Equivalently, we can define the *probability current* $J_{i,i+1}(\mu) := \mu(i)r_+ - \mu(i+1)r_- = -J_{i+1,i}(\mu)$, such that time-reversal processes are the ones for which the probability current vanishes in the steady state, $J_{i,i+1}(\pi) = 0$, and irreversible processes are the ones that possess a persistent current in the steady state: $J_{i,i+1}(\pi) \neq 0$ for some $i \in \Omega$.

We stress that in general the current $J(\mu)$ has to be interpreted as a current on the configuration space Ω, which should not be confused with a classical physical current acting on the physical domain (although the physical domain and the configuration space coincide for this example). In order to see this, we consider the more complicated situation with multiple indistinguishable (and maybe interacting) particles, where the state space can be identified with some $\Omega \subseteq \mathbb{N}_0^L$. In this case, a configuration of the system is given by $\eta = (\eta(i))_{0 \leq i < L} \in \Omega$, where $\eta(i)$ denotes the number of particles at site i. For a measures μ on Ω (representing the present distribution of the Markov chain) one then considers the probability current $J_{\eta,\eta'}(\mu) = \mu(\eta)r_{\eta,\eta'} - \mu(\eta')r_{\eta',\eta}$ between two configurations $\eta, \eta' \in \Omega$ (where $r_{\eta,\eta'}$ denotes the transition rate).

Denoting with $\eta^{i,i+1}$ the configuration obtained from η by moving one particle from site i to site $i+1$, we can construct the (averaged) physical current $\hat{\jmath}_{i,i+1}(\mu) := \sum_{\eta \in \Omega} \mu(\eta)(r_{\eta,\eta^{i,i+1}} - r_{\eta,\eta^{i+1,i}})$, which describes the expected (net) jumps of particles from site i to site $i+1$ (such that $\hat{\jmath}_{i,i+1}(\mu) = -\hat{\jmath}_{i+1,i}(\mu)$). One then can easily see that the detailed balance condition for the stationary distribution π, given by $J_{\eta,\eta^{i,i+1}}(\pi) = 0$, implies that also the physical current vanishes: $\hat{\jmath}_{i,i+1}(\pi) = 0$.

1.2.4 Time-Reversal Symmetric Processes as Gradient Flows

Due to the additional symmetry of time-reversal processes, the time-evolution can be uniquely described in terms of thermodynamic quantities, such as *entropy* and *free energy*. More precisely, under suitable assumptions on the process, one can show that the process is (for both reversible and irreversible dynamics) driven by *thermodynamic forces*, which can in the reversible case be written as gradients of thermodynamic variables. From a mathematical perspective, this allows to interpret the dynamics of the system as a *gradient flow* of the corresponding quantities. The classical example for a gradient flows [44] is the diffusion equation $\partial_t \rho_t = \Delta \rho_t$, which can formally be stated as

$$\partial_t \rho_t = \nabla \cdot \left(\rho_t \nabla \frac{\delta \mathcal{F}}{\delta \rho_t} \right)$$

for the 'energy' $\mathcal{F}(\rho) = \int \rho \log \rho \, \mathrm{d}u$. Note that the thermodynamic force is here given by the gradient $\nabla \frac{\delta \mathcal{F}}{\delta \rho} = \nabla \log \rho$. In general, such a representation is not available for irreversible systems.

1.3. Large Deviation Theory

Fluctuations in stochastic systems are intrinsically related to the theory of *large deviations*, an asymptotic theory that describes probabilities of rare events. We here give a very brief and non-rigorous overview of *Large Deviation Principles (LDPs)* and set the terminology for the different LDPs used in this thesis. For a rigorous treatment of the subject, we refer the reader to the comprehensive textbooks [11, 12, 13].

Consider a sequence of random variables $(\hat{\rho}_n)_{n\in\mathbb{N}}$ taking values in some set $\mathcal{X}$ and let $(a_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$ be a diverging and strictly monotonically increasing sequence $(a_n \nearrow \infty)$. Formally, the sequence $(\hat{\rho}_n)_{n\in\mathbb{N}}$ is said to satisfy an *LDP with speed* $(a_n)_{n\in\mathbb{N}}$ and *rate function* $\mathcal{I}\colon \mathcal{X} \to [0,\infty]$, if the probability that $\hat{\rho}_n$ is for large n 'close' to some $\rho \in \mathcal{X}$ is of order $\mathrm{e}^{-a_n \mathcal{I}(\rho)}$, which is usually denoted with

$$\mathrm{Prob}(\hat{\rho}_n \approx \rho) \asymp \mathrm{e}^{-a_n \mathcal{I}(\rho)}. \tag{1.1}$$

A mathematically more precise statement of (1.1) is to say that for any open set $O \subseteq \mathcal{X}$

$$\liminf_{n\to\infty} \frac{1}{a_n} \log \mathrm{Prob}(\hat{\rho}_n \in O) \geq - \inf_{\rho\in O} \mathcal{I}(\rho)$$

and for any closed set $A \subseteq \mathcal{X}$

$$\limsup_{n\to\infty} \frac{1}{a_n} \log \mathrm{Prob}(\hat{\rho}_n \in A) \leq - \inf_{\rho\in A} \mathcal{I}(\rho).$$

In the following, we will e.g. consider cases where the speed is given by the number of independent copies of a Markov chain $a_n = n$ or the number of lattice sites $a_L = L^d$ for an interacting particle system. A similar notation as above can be adopted in the case where the index $n \in \mathbb{N}$ is replaced by a continuous index, such as time $T \in [0,\infty)$, or the strength of a random noise $\epsilon \in [0,\infty)$, cf. [18].

In typical situations, the rate functional $\mathcal{I}$ has a unique minimum $\bar{\rho}$ with $\mathcal{I}(\bar{\rho}) = 0$ and this minimum characterises the expected behaviour of the system. For example, for a sequence of independent and real-valued random variables $(Z_i)_{i\in\mathbb{N}}$ with mean $m \in \mathbb{R}$, the rate functional associated to $\frac{1}{n}\sum_{i=1}^n Z_i$ is uniquely minimised at m, as it is predicted by the weak law of large numbers. Cramér's Theorem (see Theorem 2.2.3 in [11]) shows that a deviation from the mean can only happen with an exponentially small probability, which vanishes in the limit as $n \to \infty$. Note that this result is an improvement over the weak law of large numbers, as it does not only show that $\frac{1}{n}\sum_{i=1}^n Z_i$ converges in probability to m, but also yields an exponential rate for the concentration of the distribution.

From a physics perspective, one expects that the minimisers of the rate function yield relevant information on the system under consideration. For example, one can show that classical fundamental relations between *thermodynamic quantities* from *equilibrium thermodynamics*, such as *entropy* and *free energy*, can be recovered as scaling limits for appropriate statistical mechanics models, such as spin systems and interacting particle systems, with diverging system size. We refer the reader to e.g. Section 3.5 in [6] or [50] for a detailed discussion. We further stress that *phase transitions* in

physical systems are often related to rate functions that are not (strictly) convex, see e.g. [8, 1, 6, 13]. These ideas a priori motivate our interest in general structures for large deviation rate functions.

1.3.1 LDPs for Markov Chains

In the following chapters we will be interested in several large deviation principles for Markov chains. We consider the pathwise LDP, the LDPs at level-2.5 and level-2 and also the LDP for current fluctuations, which we now introduce briefly.

To any time-continuous Markov chain $(X_t)_{t\geq 0}$, for the sake of simplicity with a finite state space V, we can associate an *empirical density* δ_{X_t} (which is a random probability measure on V). Further, for $x, y \in V$ and $t > 0$ we denote with $C_{xy}(t)$ the *empirical flow*, another random quantity which accounts for the instantaneous jumps from x to y at time t, such that $\frac{1}{T}\int_0^T C_{xy}(t)\,\mathrm{d}t$ is the time averaged number of jumps from x to y in $[0,T]$. The empirical flow gives rise to the *empirical current* $(\jmath_t)_{xy} = C_{xy}(t) - C_{yx}(t)$, which describes the net transitions from x to y at time t (see e.g. [3] for rigorous definitions and further details). In this thesis, we will focus on empirical currents (rather than empirical flows).

Let us describe the different LDPs mentioned above: We start with the *pathwise LDP*, for which we let $(X^i)_{i\in\mathbb{N}}$ be many independent copies of the above Markov chain on a fixed time interval $[0,T]$ and denote their currents with $(\jmath^i)_{i\in\mathbb{N}}$. We further define the empirical average $\hat{\rho}_t^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_t^i}$, as well as the averaged empirical current $\hat{\jmath}_t^n := \frac{1}{n}\sum_{i=1}^n \jmath_t^i$. The deviation of the path $(\hat{\rho}_t^n, \hat{\jmath}_t^n)_{t\in[0,T]}$ from a given path $(\rho_t, \jmath_t)_{t\in[0,T]}$ in the limit as $n\to\infty$ is then characterised by the rate function $I_{[0,T]}((\rho_t,\jmath_t)_{t\in[0,T]})$ with speed n. The validity of this LDP follows from Sanov's theorem (see Section 6.2 in [11]), which implies that the rate function is given by the *relative entropy* on the space of càdlàg paths (see e.g. Section 3 in [5]).

For a single copy of the Markov chain we will also consider fluctuations of the time-averaged quantities $\big(\frac{1}{T}\int_0^T \delta_{X_t}\,\mathrm{d}t, \frac{1}{T}\int_0^T \jmath_t\,\mathrm{d}t\big)$ in the asymptotic limit as $T\to\infty$. We refer to this as an *LDP at level-2.5* (with speed T) and the rate function will be denoted with $I_{2.5}(\rho, j)$. The nomenclature level-2.5 stems from a classification of LDPs in three levels (level-1, level-2 and level-3), which goes back to Donsker and Varadhan, see e.g. [13]. The rate functions for lower levels can be obtained from the ones for higher levels by contraction (cf. [11]). Level-2.5 was originally introduced (as an intermediate level) in [31] (see also the discussion in [3]). There, the LDP was in fact defined for the empirical density and the empirical flow $\big(\frac{1}{T}\int_0^T \delta_{X_t}\,\mathrm{d}t, \frac{1}{T}\int_0^T C(t)\,\mathrm{d}t\big)$. Note that the LDP for empirical density and empirical current (which we call LDP at level-2.5) can be obtained from this LDP for the empirical density and the empirical flow by contraction.

Similarly, the *LDP at level-2*, which considers the empirical average of the density $\frac{1}{T}\int_0^T \delta_{X_t}\,\mathrm{d}t$ only, has a rate function $I_2(\rho)$ (again with speed T) which can be obtained from level-2.5 via the contraction $I_2(\rho) = \inf_j I_{2.5}(\rho, j)$. Finally, the LDP for the averaged empirical current $\frac{1}{T}\int_0^T \jmath_t\,\mathrm{d}t$ (with speed T) has the rate function $I_{\text{current}}(j) = \inf_\rho I_{2.5}(\rho, j)$.

1.3.2 LDPs on the Macroscopic Scale

We now discuss large deviations in the context of scaling limits for interacting particle systems. Consider a family of Markov chains $(\hat{\rho}^L)_{L\in\mathbb{N}}$, such that $(\hat{\rho}_t^L)_{t\geq 0}$ is a particle system on the discrete torus $\mathbb{T}_L^d = \mathbb{Z}^d/(L\mathbb{Z}^d)$ (which is a periodic lattice with L^d sites).

This domain is embedded in the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. We assume that the initial distribution for each of these particle systems is sampled from the steady state, in which case the initial distribution satisfies an LDP

$$\mathrm{Prob}\big(\hat{\rho}_0^L \approx \rho_0\big) \asymp \mathrm{e}^{-L^d \mathcal{V}(\rho_0)},$$

which describes the probability that the density $\hat{\rho}_0^L$ at time $t = 0$ is asymptotically, as $L \to \infty$, close to some mass density $\rho_0 \colon \mathbb{T}^d \to [0, \infty)$. The rate function is given by the quasipotential $\mathcal{V}$, which is a fundamental quantity in MFT [2], where it plays, as stated before, the role of a free energy for non-equilibrium systems.

For the dynamics, we consider a path $(\rho_t)_{t\in[0,T]}$ with $\rho_t \colon \mathbb{T}^d \to [0, \infty)$ and a current $(j_t)_{t\in[0,T]}$, such that the *continuity equation* $\partial_t \rho_t = -\nabla \cdot j_t$ is satisfied. The rate function associated to the fluctuation (given the initial distribution ρ_0) is then given by

$$\mathrm{Prob}\big((\hat{\rho}_t^L, \hat{\jmath}_t^L)_{t\in[0,T]} \approx (\rho_t, j_t)_{t\in[0,T]}\big) \asymp \mathrm{e}^{-L^d \mathcal{I}((\rho_t, j_t)_{t\in[0,T]})},$$

where the rate function is given by

$$\mathcal{I}((\rho_t, j_t)_{t\in[0,T]}) = \mathcal{V}(\rho_0) + \frac{1}{4} \int_0^T \int_{\mathbb{T}^d} \big(j_t - J(\rho_t)\big) \cdot \chi(\rho_t)^{-1} \big(j_t - J(\rho_t)\big) \,\mathrm{d}u \,\mathrm{d}t,$$

which depends on mobility $\chi(\rho_t(u))$ of the underlying particle system.

In the following, we will also be interested in empirical averages of the density and current, which lead to level-2.5 and level-2 like formulas for MFT. For this, we consider the empirical averages $(\frac{1}{T}\int_0^T \hat{\rho}_t^L \,\mathrm{d}t, \frac{1}{T}\int_0^T \hat{\jmath}_t^L \,\mathrm{d}t)$ and $(\frac{1}{T}\int_0^T \hat{\rho}_t^L \,\mathrm{d}t)$, with 'speed' TL^d, in the limit as first $L \to \infty$ and then $T \to \infty$, as it is done in [1].

1.4. Outline of the Thesis

Chapters 2 to 4 are the main body of this thesis. In each of these chapters we present a research paper which originated from the PhD study of the author.

Chapter 2 contains our first paper, in which we investigate the acceleration of convergence for time-reversal symmetric systems by breaking detailed balance. This question is interesting for (at least) two different reasons. First of all, from the theoretical perspective, a better understanding of how stochastic processes converge to their steady states yields new and interesting insights, which can lead to a better understanding of how to design physical systems. Further, these ideas can be used to understand and improve existing sampling algorithms, which are e.g. used in computational physics and Bayesian statistics. In this chapter, we focus on Markov chains, which in particular covers interacting particle systems, and hydrodynamic scaling limits.

In Chapter 3, we discuss a variational (or canonical) structure for (finite state) Markov chains, which allows to characterise fluctuations in the systems (described by the rate functions defined in Section 1.3). We show that these rate functions can be represented in terms of a functional called the *Onsager-Machlup functional* Φ, first introduced in [43]. This functional is non-quadratic, which in particular allows to conclude that Markov chains satisfy a non-linear flux-force relation. We discuss the relation to Ψ-$\Psi^\star$ structures, which leads to a general framework, which is valid for Markov chains, single particle diffusions and macroscopic systems (as described by MFT). In particular, we show how the *Hamilton-Jacobi orthogonality* from MFT can be generalised to a 'generalised orthogonality' which involves the $\Psi^\star$ term and the

thermodynamic force acting on the system. This generalised orthogonality allows us to derive results for Markov chains, which are in analogy to the results obtained for MFT in Section 2.2 in Chapter 2.

In Chapter 4, we show how the macroscopic description from Chapter 3 can, for certain models, be obtained in the limit of the corresponding quantities for interacting particle systems. In Chapter 3 we claim that the structure of MFT (which is defined on the physical domain) arises in the asymptotic limit from the structure of the underlying interacting particle system.

Finally, Chapter 5 contains our final conclusions and a brief outlook on possible future research directions.

Chapter 2

Acceleration of Convergence to Equilibrium in Markov Chains by Breaking Detailed Balance

This chapter is devoted to the question why (and how) irreversible processes tend to converge faster to equilibrium than reversible processes. We present joint work with Robert L. Jack and Johannes Zimmer, which was published in the *Journal of Statistical Physics* as open access publication [28].

2.1. Outline of the Article

Consider an ergodic Markov chain with unique steady state π. The dynamics of this Markov chain can be described in terms of a generator $\mathcal{L}$. If the steady state π is known (at least up to a constant), one can split $\mathcal{L}$ in a *time reversal symmetric* part $\mathcal{L}_S$ (which itself is a generator of a Markov chain) and a second part $\mathcal{L}_A$ which is anti-symmetric under time-reversal (see Section 2.1 below). Note that the Markov chain corresponding to $\mathcal{L}_S$ satisfies the detailed balance condition w.r.t. π and also converges to π in the long time limit ($t \to \infty$), just as $\mathcal{L}$ does. Alternatively, given a time-reversal symmetric generator $\mathcal{L}_S$, one may modify the transition rates in such a way that π is still the invariant measure, but the detailed balance condition is no longer valid. One then can compare the relaxation of the two processes with generator $\mathcal{L}_S$ and generator $\mathcal{L}$ to π. In order to have a 'fair' comparison, it seems reasonable to consider only processes where the symmetric part of $\mathcal{L}$ coincides with $\mathcal{L}_S$ (as multiplying the rates with a constant $c > 1$ would trivially accelerate the convergence of the process).

In the literature it is by now well known that this process of 'breaking detailed balance' (by introducing $\mathcal{L}_A$) tends to accelerate the convergence of Markov processes (such as Markov chains and stochastic differential equations (SDEs)), see e.g. [4, 22, 35, 47]. This includes in particular interacting particle systems that can be described in terms of Markov chains, like the simple exclusion process (SEP) and the zero-range process (ZRP) [32], on which we will focus in this chapter.

In Section 2.1 we review classical spectral gap and large deviation arguments for Markov chains [4]. In Section 2.2 we show how one can leverage large deviation arguments for SDEs, obtained in [46], to diffusive scaling limits of interacting particle systems [32], as described by *Macroscopic Fluctuation Theory (MFT)* [2], where the role of $\mathcal{L}$ is played by a current $J(\rho)$ (depending on a mass density ρ), which can also be split in a symmetric and anti-symmetric part, $J(\rho) = J_S(\rho) + J_A(\rho)$.

In Section 3 we support our theoretical findings with numerical simulations of independent particles and the zero-range process in one and two space dimensions. A data set with the simulation results and the MATLAB code used for the simulations was made available on the University of Bath data archive under DOI:10.15125/BATH-00365.

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
Acceleration of Convergence to Equilibrium in Markov Chains by Breaking Detailed Balance									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**		**Accepted**		**Published**	X
Publication details (reference)	Journal: Journal of Statistical Physics, DOI: 10.1007/s10955-017-1805-z Authors: Marcus Kaiser, Robert L. Jack, Johannes Zimmer								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The bulk of the calculations have been performed by the author of the thesis (70%). All authors contributed equally to the presentation of the content (33%). The numerical computations have been performed by the author of the thesis (100%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed					**Date**	10.7.2018			

J Stat Phys (2017) 168:259–287
DOI 10.1007/s10955-017-1805-z

Acceleration of Convergence to Equilibrium in Markov Chains by Breaking Detailed Balance

Marcus Kaiser[1] · **Robert L. Jack**[2] · **Johannes Zimmer**[1]

Received: 22 November 2016 / Accepted: 2 May 2017 / Published online: 18 May 2017

Abstract We analyse and interpret the effects of breaking detailed balance on the convergence to equilibrium of conservative interacting particle systems and their hydrodynamic scaling limits. For finite systems of interacting particles, we review existing results showing that irreversible processes converge faster to their steady state than reversible ones. We show how this behaviour appears in the hydrodynamic limit of such processes, as described by macroscopic fluctuation theory, and we provide a quantitative expression for the acceleration of convergence in this setting. We give a geometrical interpretation of this acceleration, in terms of currents that are *antisymmetric* under time-reversal and orthogonal to the free energy gradient, which act to drive the system away from states where (reversible) gradient-descent dynamics result in slow convergence to equilibrium.

Keywords Convergence to equilibrium · Non-equilibrium processes · Zero-range process · Macroscopic Fluctuation Theory · Large deviations

1 Introduction

In this paper we analyse the effects of breaking detailed balance for interacting particle systems (as described by Markov processes [32]), and their hydrodynamic scaling limits (as described by Macroscopic Fluctuation Theory [9]). The interacting particle systems represent microscopic descriptions of physical systems, in which the motion of each particle may be

Marcus Kaiser
m.kaiser@bath.ac.uk

Robert L. Jack
r.jack@bath.ac.uk

Johannes Zimmer
j.zimmer@bath.ac.uk

[1] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

[2] Department of Physics, University of Bath, Bath BA2 7AY, UK

followed individually. The (fluctuating) hydrodynamic model of the same system describes its behaviour on large length and time scales, in which case the motion of the individual particles is no longer visible, and one works instead with a smooth density field, whose time evolution includes a deterministic element as well as a (weak) stochastic noise [28].

Among interacting particle systems, those with detailed balance are special—they correspond to Markov chains that are reversible with respect to an invariant measure π. Physically, these models are important because their steady states are time-reversal symmetric and lack any persistent currents, so they can be used to describe systems that relax to states of thermal equilibrium. They also have applications outside physics, because given a (possibly non-normalised) measure ν, it is straightforward to design a reversible Markov chain whose invariant measure π is proportional to ν. This construction is at the root of many Markov chain Monte Carlo (MCMC) methods [2,34], in which one typically aims to generate large numbers of uncorrelated samples from a prescribed distribution π. Such methods have widespread applications including Bayesian learning, protein folding and cryptography [14].

In both the physical systems and the MCMC methods, an important question is the rate of convergence to equilibrium of the relevant Markov chains. In MCMC, this rate controls the computational cost required to obtain independent samples from π, which is an important factor in the efficiency of the method. In the physical systems, the question of how fast a system converges to equilibrium controls many physical properties including fluid viscosities, and systems' abilities to respond to changes in external conditions, such as temperature.

Recently, several results have become available which show that for a given invariant measure π, reversible Markov chains have the slowest convergence [10,24,30,35,36,39]. Given that most common MCMC methods are based on such reversible models, and that faster convergence is linked to improved efficiency, this observation offers a route towards the development of new and more efficient methods, some of which are already becoming available [5]. Breaking reversibility can be achieved by an explicit modification of transition rates [36], or by an expansion of the state space (*lifting*) to incorporate persistence of motion or inertial effects [12,15]. The main physical feature of the resulting irreversible Markov chains is that they (generically) have non-equilibrium steady states characterised by finite entropy production and dissipation of energy. Compared to the equilibrium setting, the nature of fluctuations and convergence to steady state in non-equilibrium systems is much less understood, and is an area of important current activity [3,9,13].

To address these questions, this paper presents several new results. First, we revisit existing results for microscopic models, concentrating in particular on the spectral gap of the generator, and how it is affected when detailed balance is broken. Second, we investigate how breaking detailed balance affects the hydrodynamic limit of the model—in this latter case, convergence to equilibrium is most easily analysed via large deviations of the empirical measure [35,36]. Third, we illustrate our general results by numerical results of a simple interacting particle system—the zero-range process [38]. These numerical results are particularly relevant since the analytical results indicate that breaking detailed balance can never slow down convergence to equilibrium, but they provide rather little insight into how much this convergence can be accelerated, nor how this effect depends on the specific way in which detailed balance is broken. We provide some general remarks and comments in this direction.

1.1 Characterisation of Convergence to Steady State

A number of methods are available to analyse the time required for a system to reach its steady state. This section contains a brief review of some of them. For microscopic models—such as Markov processes on (finite) discrete spaces and SDEs—we mention some recent

work showing how breaking detailed balance can accelerate convergence of systems to their steady states. These results serve as a foundation for our results here, which show how these effects manifest on the macroscopic scale.

1.1.1 Spectral Gap

The first—and most common—method for analysis of convergence to equilibrium is to estimate the spectral gap of the generator of the relevant stochastic process. In general, the eigenvalues $\{\lambda_i\}$ of the generator are complex numbers, there is a simple eigenvalue $\lambda_0 = 0$ and all other eigenvalues have negative real parts. The spectral gap $\alpha_{\min}$ is the minimal value of $|\lambda_i^r|$ among the non-zero eigenvalues, where λ_i^r denotes the real part of λ_i. Roughly speaking, the physical significance of the spectral gap is that the system converges exponentially fast to its steady state, with a characteristic time scale

$$\tau_{\mathrm{g}} = (1/\alpha_{\min}). \tag{1}$$

For stochastic differential equations [24,30] and discrete-space Markov processes [36], it has been shown that irreversible processes generically have smaller time scales τ_{g}, compared to reversible processes with the same invariant measure. We provide further results in this direction in Sect. 2.1.1 below, for the discrete space Markov processes that are relevant for interacting particle systems.

1.1.2 Asymptotic Variance

Another set of methods for the analysis of the convergence to steady state is based on empirical time averages. That is, let X_t be the state of the system at time t and let f be an observable quantity (test function) whose value at time t is $f(X_t)$. Then the empirical time average of f is

$$\overline{f}(T) := \frac{1}{T}\int_0^T f(X_s)\,ds. \tag{2}$$

The quantity $\overline{f}(T)$ is a random variable which—under suitable conditions related to ergodicity—converges almost surely to the expectation value of f, which we denote by $\mathbb{E}_\pi(f)$.

Moreover the distribution of $\sqrt{T}(\bar{f}(T)-\mathbb{E}_\pi(f))$ converges by the central limit theorem to a normal distribution with variance σ_f^2. The latter is referred to as asymptotic variance or time average variance constant (TAVC) which can be obtained as $\sigma_f^2 = \lim_{T\to\infty} T\mathrm{Var}(\bar{f}(T))$, see in [2, Chapter IV], [35] and [40, Sect. 3.5]. Hence, the variance of $\overline{f}(T)$ decays for large times as $\mathrm{Var}(\overline{f}(T)) \sim \sigma_f^2/T$. It is then natural to identify a time scale $\tau_{\mathrm{v}}^f := \sigma_f^2/\mathrm{Var}_\pi(f)$. Note that τ_{v}^f depends on the observable f of interest—roughly speaking it represents the autocorrelation time of $f(X_t)$. In general τ_{v}^f and τ_{g} are different time scales: τ_{v}^f controls the convergence of $\overline{f}(T)$ while τ_{g} controls the convergence of the probability measure itself. As for τ_{g}, one finds that σ_f^2 can be reduced by breaking detailed balance in Markov chains [10,39] and SDEs [18,25].

1.1.3 Large Deviations at Level-1 and Level-2

A more detailed analysis of the large-T behaviour of $\overline{f}(T)$ is available from large deviation theory [23,40]. Informally, one expects that for large T, the random variable $\overline{f}(T)$ satisfies

$$\text{Prob}\left[\overline{f}(T) \approx \hat{f}\right] \asymp \mathrm{e}^{-T I_f(\hat{f})} \tag{3}$$

for some *rate function* I_f (which depends on the choice of test function f). We use the notation in (3) throughout this work as an informal way to state large deviation principles: it means that the log probability that $\overline{f}(T)$ takes a value in a small interval containing $\hat{f}$ can be bounded above and below by quantities related to the rate function I_f [23,40]. The rate function achieves its minimal value of zero when $\hat{f}$ is equal to $\mathbb{E}_\pi(f)$, and the second derivative of I_f at this minimum is related to σ_f^2. The function I_f is a level-1 rate function [23].

A yet more detailed analysis is available by considering not just the large deviations of a single test function f but instead to consider large deviations of the empirical measure. That is, for a Markov chain on a discrete space Ω, define the empirical measure

$$\bar{\mu}_T(x) := \frac{1}{T}\int_0^T \delta_{X_s,x}\, ds, \tag{4}$$

where $\delta_{x,y}$ is a Kronecker delta. The empirical measure at time t is a vector $\bar{\mu}_T = (\bar{\mu}_T(x))_{x\in\Omega}$. For large enough T, ergodicity implies that $\bar{\mu}_T(x)$ converges almost surely to $\pi(x)$, and the fluctuations of the measure μ in this limit are described by a large deviation principle at level-2:

$$\text{Prob}\left[\bar{\mu}_T \approx \nu\right] \asymp \mathrm{e}^{-T I_2(\nu)} \tag{5}$$

where I_2 is the rate function [17], which now depends on a vector ν instead of a single real argument $\hat{f}$. Note that the (level-1) rate function I_f for any observable f can be obtained by a contraction of this large deviation principle, so the function I_2 contains a great deal of information about the convergence of a system to its steady state. Moreover, as might be expected from the terminology "rate function", the quantity $1/I_2(\mu)$ has an interpretation as a μ-dependent time scale associated with the decay of an initial measure μ to the invariant measure π.

Recent work by Rey-Bellet and Spiliopoulos [35,36] has motivated the analysis of I_2 as a measure of the rate of convergence of processes to their steady states. Their work, and that of Bierkens [10], show that breaking detailed balance accelerates this convergence. Note however that in contrast to the spectral gap—where a single number characterises the rate of convergence of the whole system—the rate function I_2 depends on the measure μ for which it is evaluated; similarly the asymptotic variance σ_f^2 depends on the specific observable f. In this sense, the information available from the asymptotic variance and the large deviations is greater than that available from the spectral gap, but this extra information may also make these measures harder to interpret in terms of simple acceleration or slowing down of convergence to equilibrium. Of course, other useful measurements of convergence rates are available, such as mixing times [31], cutoff phenomena (see e.g. [29], where cutoff was recently established for the asymmetric simple exclusion process) and log-Sobolev constants (e.g. [16]), but these are not analysed in this work.

1.2 Outline

The remainder of this paper is organised as follows: Sect. 2 includes a theoretical analysis of the effects of breaking detailed balance on convergence to steady states, including both Markov chains (Sect. 2.1) and hydrodynamic limits (Sect. 2.2). Section 3 presents numerical results that illustrate this acceleration in the zero-range process: we provides examples in both one-dimensional and two-dimensional settings. Finally, Sect. 4 contains our conclusions.

2 Theoretical Results

2.1 Acceleration of the Microscopic Dynamics

In this section, we consider an irreducible Markov jump process on a finite state space Ω which contains n states. In terms of interacting particle systems, this process describes the dynamics of a finite number of particles that move on some finite lattice. The process is defined by rates $c(x \to y)$ for states $x, y \in \Omega$. The condition of *detailed balance* (or *reversibility*) is that for some probability measure π and all x, y then

$$\pi(x)c(x \to y) = \pi(y)c(y \to x). \tag{6}$$

In this case the (unique) invariant measure of the Markov process is π.

Let the generator of the Markov process be $\mathcal{L}$. The generator has a representation as an $n \times n$ matrix and the reversibility condition (6) corresponds to symmetry of $\mathcal{L}$ with respect to the $L^2(\pi)$ inner product $\langle f, g\rangle_\pi = \sum_x f(x)g(x)\pi(x)$. If detailed balance is broken (non-reversible Markov chain), then $\mathcal{L}$ is not symmetric with respect to $L^2(\pi)$, but one may always write

$$\mathcal{L} = \mathcal{L}_S + \mathcal{L}_A, \tag{7}$$

where $\mathcal{L}_S$ is symmetric with respect to $L^2(\pi)$, while $\mathcal{L}_A$ is antisymmetric. Moreover, $\mathcal{L}_S$ is a generator for a reversible stochastic process, whose transition rates may be verified to be

$$c_s(x \to y) = \frac{1}{2}\left[c(x \to y) + \pi(y)c(y \to x)\pi(x)^{-1}\right], \tag{8}$$

where π is the invariant measure of $\mathcal{L}$, which is also the invariant measure of $\mathcal{L}_S$. (Recall that the original Markov process is finite and irreducible, which ensures that $\pi(x) > 0$ for all x). We also identify the off-diagonal elements of $\mathcal{L}_A$ as

$$c_a(x \to y) := c(x \to y) - c_s(x \to y).$$

Hence one has from (8) that

$$\pi(x)[c_s(x \to y) + c_a(x \to y)] = \pi(y)[c_s(y \to x) - c_a(y \to x)]. \tag{9}$$

Note that $\mathcal{L}$ and $\mathcal{L}_S$ both are generators, whereas the operator $\mathcal{L}_A$ is not a generator of a Markov chain.

Alternatively one can think of the decomposition of $\mathcal{L}$ in $\mathcal{L}_S$ and $\mathcal{L}_A$ as follows: consider the Markov process η_t (with $t \in [-T, T]$ for some $T > 0$) associated to $\mathcal{L}$, with the initial condition distributed by the steady state π. The time reversed process $\hat{\eta}(t) := \eta(-t)$ is also associated to a generator, $\mathcal{L}^*$, say. The symmetric part of the generator can be recovered as $\mathcal{L}_S = (\mathcal{L} + \mathcal{L}^*)/2$.

Given these preliminaries, we can now be precise about the sense in which breaking detailed balance accelerates convergence: in all cases we compare the process $\mathcal{L}$ with the corresponding symmetrised process $\mathcal{L}_S$. (Equivalently, one may imagine starting from a reversible process $\mathcal{L}_S$ and breaking detailed balance by adding an extra term $\mathcal{L}_A$ to the generator.) The processes $\mathcal{L}$ and $\mathcal{L}_S$ both converge to the same invariant measure π—one aims to prove that convergence times such as τ_g or $1/I(\mu)$ are smaller for $\mathcal{L}$ than for $\mathcal{L}_S$.

2.1.1 The Spectral Gap

To illustrate how breaking detailed balance accelerates convergence, we show in Proposition 1 below that breaking detailed balance can only increase the spectral gap, so that the convergence of the irreversible process is characterised by a smaller value of the time τ_g. This result has been proven in greater generality in [24,36], but we provide a short proof here, for illustrative purposes.

To this end, consider an initial measure μ_0, and represent it in terms of an eigendecomposition of $\mathcal{L}$, so that

$$\mu_0(x) = \pi(x) + \sum_{j=1}^{m} \big(\alpha_j \nu_j(x) + \bar{\alpha}_j \bar{\nu}_j(x)\big), \tag{10}$$

where the $\alpha_j \in \mathbb{C}$ are μ_0-dependent coefficients, while ν_j are complex-valued measures which are left-eigenvectors corresponding to eigenvalues λ_j of $\mathcal{L}$. The overbar (e.g. $\bar{\alpha}$) denotes the complex conjugate. Decomposing the non-zero eigenvalues λ_j into real and imaginary parts, as $\lambda_j = \lambda_j^r + i\lambda_j^i$, the measure at time t is given by

$$\mu_t = (\mathrm{e}^{t\mathcal{L}})^\dagger \mu_0 = \pi + \sum_{j=1}^{m} \mathrm{e}^{\lambda_j^r t}\big(\mathrm{e}^{i\lambda_j^i t}\alpha_j \nu_j + \mathrm{e}^{-i\lambda_j^i t}\bar{\alpha}_j \bar{\nu}_j\big), \tag{11}$$

where $\cdot^\dagger$ denotes a matrix transpose. Note that for real-valued eigenvalues (with $\lambda_j^i = 0$) the term in brackets is equal to $2\alpha_j \nu_j$, as in this case also the left (and right) eigenvectors are real-valued.

Moreover, $\lambda_j^r < 0$ for all j, since $\mathcal{L}$ is the generator of an irreducible finite Markov process. One sees immediately that this Markov process relaxes exponentially fast to its steady state. Moreover, the rate of this exponential decay is controlled by the non-zero eigenvalue of $\mathcal{L}$ whose real part is smallest in magnitude. Similar results to the following proposition have already been obtained in e.g. [26,37]:

Proposition 1 *Let $\mathcal{L}$ and $\mathcal{L}_S$ be given as above. The non-zero eigenvalues of $-\mathcal{L}_S$ are real and positive; let the smallest such eigenvalue be $\alpha_{\min}$ and the largest be $\alpha_{\max}$. Then every non-zero eigenvalue λ of $-\mathcal{L}$ satisfies*

$$\alpha_{\min} \leq \mathrm{Re}(\lambda) \leq \alpha_{\max}. \tag{12}$$

Proof Define the Dirichlet form for $\mathcal{L}$ as $\mathcal{E}(f, g) := \langle f, -\mathcal{L}g\rangle_\pi = \sum_x f(x)(-\mathcal{L}g)(x)\pi(x)$, where π is the unique stationary distribution of $\mathcal{L}$. Let λ be a non-zero eigenvalue of $-\mathcal{L}$ with corresponding right eigenvector $f_\lambda + ig_\lambda$. As $\mathcal{E}(1, f) = 0$ for all f, we obtain $0 = \mathcal{E}(1, f_\lambda + ig_\lambda) = \lambda(\langle 1, f_\lambda\rangle_\pi + i\langle 1, g_\lambda\rangle_\pi)$. Since λ is non-zero, we obtain that $\langle 1, f_\lambda\rangle_\pi = 0 = \langle 1, g_\lambda\rangle_\pi$. This implies that both f_λ and g_λ are mean zero, so $\mathrm{Var}_\pi(h) = \langle h, h\rangle_\pi$ for $h \in \{f_\lambda, g_\lambda\}$. Since $\mathcal{E}(f_\lambda - ig_\lambda, f_\lambda + ig_\lambda) = \lambda\langle f_\lambda - ig_\lambda, f_\lambda + ig_\lambda\rangle_\pi = \lambda(\langle f_\lambda, f_\lambda\rangle_\pi + \langle g_\lambda, g_\lambda\rangle_\pi)$, the bilinearity of the Dirichlet form yields that the real part of λ is given by

$$\mathrm{Re}(\lambda) = \frac{\mathcal{E}(f_\lambda, f_\lambda) + \mathcal{E}(g_\lambda, g_\lambda)}{\langle f_\lambda, f_\lambda\rangle_\pi + \langle g_\lambda, g_\lambda\rangle_\pi}. \tag{13}$$

In addition, one has (for the min and max taken over the two cases $h = f_\lambda$ and $h = g_\lambda$)

$$\min_{h\in\{f_\lambda, g_\lambda\}} \frac{\mathcal{E}(h, h)}{\langle h, h\rangle_\pi} \leq \frac{\mathcal{E}(f_\lambda, f_\lambda) + \mathcal{E}(g_\lambda, g_\lambda)}{\langle f_\lambda, f_\lambda\rangle_\pi + \langle g_\lambda, g_\lambda\rangle_\pi} \leq \max_{h\in\{f_\lambda, g_\lambda\}} \frac{\mathcal{E}(h, h)}{\langle h, h\rangle_\pi}. \tag{14}$$

Define $\mathcal{E}_S(f,g) = \langle f, -\mathcal{L}_S g\rangle_\pi$, and note that $\mathcal{E}(h,h) = \mathcal{E}_S(h,h)$. Also $\alpha_{\min} = \min_{h:\langle 1,h\rangle_\pi=0} \frac{\mathcal{E}_S(h,h)}{\langle h,h\rangle_\pi}$. Hence the left hand side of (14) is bounded below by $\alpha_{\min}$. Applying a similar argument to the right hand side of (14) and combining with (13) finally yields (12). □

2.1.2 Bounds on Level-2 Rate Functions for Discrete Markov Processes

From Proposition 1 and using (1), one clearly has

$$\tau_{\mathrm{g}}^{\mathrm{irr}} \le \tau_{\mathrm{g}}^{\mathrm{rev}}. \tag{15}$$

That is, the irreversible process converges to its steady state at least as quickly as the reversible one. A similar argument [10] establishes that the level-2 rate functions for $\mathcal{L}$ and $\mathcal{L}_S$ are related as

$$I_2(\mu) \ge I_2^S(\mu), \tag{16}$$

again establishing a faster rate of convergence on breaking detailed balance. Recall that results of the form (16) yield information about the empirical measure $\bar{\mu}_T$ defined in (4), whereas the previous result (12) concerns the spectral gap and the convergence of μ_t, the distribution of the process at time t as defined in (11). Note that $\bar{\mu}_T$ is a random quantity, whereas μ_t is the solution to a deterministic differential equation.

We now show (Proposition 2) that the rate of convergence of the irreversible model has an upper bound, as well as the lower bound given by $I_2^S(\mu)$. That is, $I_2(\mu)$ is bounded both above and below, just as the spectral gap is bounded in (12). This limits the acceleration that is available by breaking detailed balance for (finite) discrete Markov processes, in contrast to the situation for diffusions [35]. The proof for the following proposition is based on the variational formula for the level-2 LDP [23]. Whilst the lower bound, which is known in the literature, see e.g. [10,36], follows from the variational representation of the rate function, the upper bound is (to our knowledge) a novel result.

Proposition 2 *Consider a finite-state continuous-time Markov chain with generator $\mathcal{L} = \mathcal{L}_S + \mathcal{L}_A$ and transition rates $c(x \to y) = c_s(x \to y) + c_a(x \to y)$, as defined in Sect. 2.1. The level-2 rate functional $I_2(\mu)$ is bounded as follows:*

$$I_2^S(\mu) \le I_2(\mu) \le I_2^S(\mu) + \sum_{x\neq y}\left[c_s(x\to y) - \sqrt{c_s(x\to y)^2 - c_a(x\to y)^2}\right]\sqrt{\tfrac{\mu(x)}{\pi(x)}\tfrac{\mu(y)}{\pi(y)}}\pi(x), \tag{17}$$

where the rate functional $I_2^S(\mu)$ for the reversible process with generator $\mathcal{L}_S$ is given by $I_2^S(\mu) = \langle\sqrt{\frac{\mu}{\pi}}, -\mathcal{L}_S\sqrt{\frac{\mu}{\pi}}\rangle_\pi$.

Proof The rate functional is given by a variational formula [23]:

$$I_2(\mu) = \sup_{f>0}\langle f^{-1}, -\mathcal{L}f\rangle_\mu.$$

In the symmetric case ($\mathcal{L} = \mathcal{L}_S$) the maximum is $I_2^S(\mu)$, which is attained when $f = \sqrt{\mu/\pi}$. In general we write $f = \sqrt{\mu/\pi}\,\mathrm{e}^V$ for some potential V.

A direct computation yields

$$I_2^S(\mu) = \sum_{x\neq y}\left(\sqrt{\frac{\mu(x)}{\pi(x)}} - \sqrt{\frac{\mu(y)}{\pi(y)}}\right)\sqrt{\mu(x)\pi(x)}c(x\to y) \tag{18}$$

and

$$I_2(\mu) = I_2^S(\mu) + \sup_V I_A(\mu, V) \tag{19}$$

with

$$I_A(\mu, V) = \sum_{x \neq y} \sqrt{\frac{\mu(y)}{\pi(y)}} \left(1 - \mathrm{e}^{V(y)-V(x)}\right) \sqrt{\mu(x)\pi(x)} c(x \to y). \tag{20}$$

If V is a constant function, then $I_A(\mu, V) = 0$ so clearly $\sup_V I_A(\mu, V) \geq 0$. Hence, (19) yields the lower bound in (17), as in [10].

For the upper bound, it is convenient to define $m(x, y) := \frac{1}{2}\sqrt{\frac{\mu(x)\mu(y)}{\pi(x)\pi(y)}}$ and $q(x, y) := \pi(x)c(x \to y)$. This yields

$$I_A(\mu, V) = \sum_{x \neq y} m(x, y) \left[(1 - \mathrm{e}^{V(y)-V(x)})q(x, y) + (1 - \mathrm{e}^{V(x)-V(y)})q(y, x)\right], \tag{21}$$

where we have symmetrised the summand with respect to x, y. For positive constants a, b, one may easily establish the general inequality $a\mathrm{e}^V + b\mathrm{e}^{-V} \geq 2\sqrt{ab}$. Applying this inequality to the summand in (21) yields

$$I_A(\mu, V) \leq \sum_{x \neq y} m(x, y) \left[q(x, y) + q(y, x) - 2\sqrt{q(x, y)q(y, x)}\right]. \tag{22}$$

From (8), (9) one has $q(x, y) + q(y, x) = 2c_s(x \to y)\pi(x)$ and $q(x, y)q(y, x) = [c_s^2(x \to y) - c_a^2(x \to y)]\pi(x)^2$; substituting these results into (22) yields

$$I_A(\mu, V) \leq \sum_{x \neq y} \sqrt{\tfrac{\mu(x)}{\pi(x)}\tfrac{\mu(y)}{\pi(y)}} \left[c_s(x \to y) - \sqrt{c_s(x \to y)^2 - c_a(x \to y)^2}\right]\pi(x),$$

and the combination with (19) establishes the upper bound in (17). □

2.1.3 Discussion

Our intuition for the (bounded) acceleration by breaking detailed balance is as follows: for reversible processes we can think of μ_t (the distribution of the process at time t) undergoing a steepest descent process (gradient flow) for the free energy $F(t) = \sum_x \mu_t(x) \log(\mu_t(x)/\pi(x))$, within a particular geometric setting [33]. The precise nature of this geometry is immaterial for this discussion: the key point is that relaxation to equilibrium is fast when the free energy gradient is steep, and tends to be slow when it is shallow. On breaking detailed balance, the free energy still decreases monotonically, but its motion is no longer restricted to the direction of steepest descent. This can have several possible effects and the rate of change of $F(t)$ may either increase or decrease on breaking detailed balance. However, we argue that an important contribution to the acceleration of convergence arises because the irreversible component of the dynamics drives the system away from regions where the free energy gradient is shallow and into regions where it is steeper. We will demonstrate this effect explicitly at the hydrodynamic level, in Sect. 2.2.3.

Notice however, that while slow processes associated with $\mathcal{L}_S$ are accelerated by breaking detailed balance, the inequality involving $\alpha_{\max}$ in Proposition 1 implies that fast aspects of the relaxation tend to be slowed down. Indeed, $\mathrm{tr}(\mathcal{L}_A) = 0$ so $\mathrm{tr}(\mathcal{L}) = \mathrm{tr}(\mathcal{L}_S)$: since the trace is equal to the sum of the eigenvalues, one sees that if some (slow) processes are accelerated by

breaking detailed balance another set of (faster) processes must be slowed down by a similar amount. Within the intuitive picture, our interpretation is that the irreversible component of the dynamics acts to push the system away from regions where the free energy gradient is very steep, so the differences between very fast and very slow processes tend to be smoothed out by the irreversibility.

2.2 Accelerating Macroscopic Processes

In this section we consider hydrodynamic limits of interacting particle systems, as described by the macroscopic fluctuation theory (MFT) [9]. We will demonstrate that the large deviation result (16) has a counterpart at the hydrodynamic level. We also explore the geometrical interpretation of this result, and we connect our result to earlier work related to SDEs that describe the motion of single particles [35].

2.2.1 Macroscopic Fluctuation Theory

We first recall the core parts of the macroscopic fluctuation theory (MFT). For a detailed review we refer to [9]. Let $\Lambda \subseteq \mathbb{R}^d$ be a connected domain with boundary $\partial\Lambda$. For simplicity, we choose here the domain $\Lambda = [0, 1]^d$. If we consider a microscopic particle process (indexed by L), its description within MFT involves two random fields, the empirical particle density ρ_t^L and the empirical current j_t^L. Roughly speaking, for $x \in \Lambda$ then ρ_t^L is the local particle density and j_t^L is a vector that indicates the rate of particle flow.

The idea of the hydrodynamic limit is that if we observe an interacting particle system on suitably large scales of length and time, then the system can be described in terms of sufficiently smooth fields ρ and j, instead of requiring a microscopic description in which all particle positions are taken into account. The deterministic quantities ρ and j are then related by a continuity equation given by

$$\partial_t \rho_t + \nabla \cdot j_t = 0. \tag{23}$$

The domain Λ is fixed in the hydrodynamic limit. The relevance to large length and time scales in the microscopic model is that one considers a large number of particles N within a domain Λ_L of linear size L. One takes N, L to infinity together for a fixed density $\tilde{\rho}_0 = N/L^d$. The domain Λ is obtained by rescaling the (increasingly large) domain Λ_L, so that Λ remains fixed as $L \to \infty$.

Within this hydrodynamic limit, the behaviour of the system on suitably large scales of space and time becomes increasingly deterministic. For example, given a time interval $[0, T]$ and initial and final densities ρ_0 and ρ_T, the probability measure for paths connecting these initial and final states concentrates (in the hydrodynamic limit) on a single most likely path. This result can be expressed as a large deviation principle for paths, which can, following [9], be written as

$$\text{Prob}\left[(\rho_t^L, j_t^L)_{t\in[0,T]} \approx (\rho_t, j_t)_{t\in[0,T]}\right] \asymp \mathrm{e}^{-L^d \mathcal{I}(\rho, j)} \tag{24}$$

with

$$\mathcal{I}(\rho, j) = \frac{1}{4}\int_0^T \int_\Lambda (j_t - J(\rho_t)) \cdot \chi(\rho_t)^{-1}(j_t - J(\rho_t))\, dx\, dt \tag{25}$$

whenever $\partial_t \rho_t = -\nabla \cdot j_t$ is satisfied, and $\mathcal{I}(\rho, j) = \infty$ otherwise. We refer the reader to the review [9] for details on the validity of (24) for a large class of particle systems including the symmetric exclusion process and zero-range processes [28,38].

Note that in contrast to the large deviation principle in Sect. 1.1.3 which is concerned with large times, this principle involves a limit of large L, with a fixed time interval $[0, T]$.

Physically, we interpret $J(\rho_t)$ in (25) as the most likely current field j_t, given that the system has density ρ_t. Within MFT, the current is assumed [9, Eq. (2.6)] to have the form

$$J(\rho) = -D(\rho)\nabla\rho + \chi(\rho)E, \tag{26}$$

where $\chi(\rho)$ and $D(\rho)$ are symmetric positive definite $d \times d$ matrices that depend on the local density ρ, and E is a fixed (x-dependent) vector field.

Physically, D and χ correspond to a density-dependent diffusivity and mobility, while E corresponds to an external force. For a given interacting particle system, the parameters D, χ and E can (in principle) be derived from the microscopic rules of the model. These parameters (along with appropriate boundary conditions associated with $\partial\Lambda$) fully specify the rate function (25) and they fully describe the hydrodynamic limit of the interacting particle system. To fix the ideas precisely, it may be useful to note that $J(\rho)$ in (26) is itself a field, whose value at position $x \in \Lambda$ is $J(\rho)(x) = -D(\rho(x))\nabla\rho(x) + \chi(\rho(x))E(x)$.

Since $J(\rho)$ is the most likely current for a given density ρ, it follows that for a given initial condition, the path measure is dominated by paths $(\rho_t)_{t\in[0,T]}$ which solve $\partial_t\rho = -\nabla\cdot J(\rho)$. These paths have $\mathcal{I} = 0$ and are said to satisfy the hydrodynamics.

As well as the large-deviation principle for paths (24), the MFT also provides a large-deviation principle for the fluctuations of the instantaneous density, in the steady state of the system. That is, if the time T is large enough that the system has converged to its steady state, one has

$$\text{Prob}[\rho_T^L \approx \rho] \asymp \mathrm{e}^{-L^d\mathcal{V}(\rho)}, \tag{27}$$

where $\mathcal{V}$ is called the quasipotential: it determines the probability of fluctuations in the density. Eq. (27) is derived under the assumption that the adjoint dynamics satisfy a further Large Deviation principle for a rate functional $\mathcal{I}^*$. We refer to chapter II in [9] for a detailed discussion.

We assume throughout that our system has a unique steady state, for which the most likely (x-dependent) density is $\overline{\rho}$. In this case $\mathcal{V}(\overline{\rho}) = 0$ and $\mathcal{V}(\rho) > 0$ for all $\rho \neq \overline{\rho}$.

2.2.2 Reversible and Irreversible Systems

For the microscopic dynamics, we already observed that the detailed balance condition (6) describes an important special case. By starting from this case, the generator was decomposed into two components (7), corresponding to a reversible process and a correction term that captures the irreversibility. At the hydrodynamic level, there is a corresponding decomposition which takes place at the level of the current: one writes

$$J = J_S + J_A. \tag{28}$$

The symmetric part of the current is defined [9, Equ. (2.19)] as

$$J_S(\rho) = -\chi(\rho)\nabla\frac{\delta\mathcal{V}}{\delta\rho}, \tag{29}$$

where $\frac{\delta\mathcal{V}}{\delta\rho}$ denotes the functional derivative of the quasipotential introduced in Eq. (27). The antisymmetric part of the current is orthogonal to J_S, in the sense that

$$\int_\Lambda J_A(\rho)\cdot\chi^{-1}(\rho)J_S(\rho)\,dx = 0, \tag{30}$$

which is sometimes referred to as a *Hamilton-Jacobi equation*. Note that this is an orthogonality in the space of fields: the presence of the integral implies that the currents J_S and J_A do not have to be orthogonal at any specific point x. We note that on combining (29) and (30), one has $\int_\Lambda J_A(\rho) \cdot \nabla \frac{\delta \mathcal{V}}{\delta \rho} dx = 0$; integrating by parts and using (28) one sees that

$$\partial_t \mathcal{V} = \langle \partial_t \rho, \frac{\delta \mathcal{V}}{\delta \rho} \rangle = \langle \mathrm{div} J, -\frac{\delta \mathcal{V}}{\delta \rho} \rangle = -\langle J_S, \chi^{-1} J_S \rangle \tag{31}$$

which is independent of J_A. Hence the quasipotential is non-increasing for paths satisfying the hydrodynamics, and (for any given ρ_t) its time derivative is independent of J_A.

The special case in which the microscopic model is reversible has two implications for the hydrodynamic limit as described by MFT. First, reversible models lead to $J_A = 0$, so $J = J_S$. Second, assuming that correlations in the particle model occur only on the microscopic scale, the quasipotential within the MFT takes the simple (local) form [9, Eq. (2.25)]

$$\mathcal{V}(\rho) = \int_\Lambda \Big[f(\rho) - f(\overline{\rho}) - f'(\overline{\rho})(\rho - \overline{\rho}) \Big] dx, \tag{32}$$

where $f(\rho)$ is the free energy per unit volume. (The dependence of f on ρ is fixed by the microscopic model of interest; note also that both ρ and $\overline{\rho}$ depend in general on the position x, but f is a local function $f(\rho)(x) = f(\rho(x))$).

Hence for reversible microscopic models, the hydrodynamic current obeys

$$J(\rho) = J_S(\rho) = -\chi(\rho) f''(\rho) \nabla \rho + \chi(\rho) \nabla f'(\overline{\rho}). \tag{33}$$

In this case consistency with (26) requires

$$E = \nabla f'(\overline{\rho}), \qquad D(\rho) = f''(\rho) \chi(\rho). \tag{34}$$

The second of these conditions is required within MFT. It is known as the local Einstein relation since it relates the mobility χ to the diffusion constant D. Note that the equations (34) are consistent with the hydrodynamic limit for a large class of particle systems of 'gradient type', see [9, Chap. VIII, Sect. G].

We end this section with a brief comment on the boundary conditions within MFT. If the boundary is associated with coupling of the system to a reservoir at chemical potential λ, the density at the boundary is fixed such that $f'(\rho) = \lambda$. If particles cannot penetrate the boundaries, one requires $D \nabla \rho = \chi E$ (and $j = 0$) on $\partial \Lambda$. Paths (or configurations) that do not respect these boundary conditions have $\mathcal{I} = \infty$.

2.2.3 Breaking Detailed Balance Accelerates Convergence

We now state the sense in which breaking detailed balance accelerates convergence of interacting particle systems at the hydrodynamic scale. For the microscopic models, we compared two Markov chains, with the same invariant measure and generators $\mathcal{L}$ and $\mathcal{L}_S$. At the hydrodynamic scale, we will compare two systems with the same quasipotential (this corresponds to comparing two microscopic models with the same invariant measure). One system is irreversible and has a general J given by (26); the second system is reversible and so $J_A = 0$. In order to ensure a fair comparison, we also assume that the two models have the same mobility $\chi(\rho)$: for Markov processes the equivalent condition was that we always compared models with the same $\mathcal{L}_S$. Since $\mathcal{V}$ and χ are the same for both models, they both have the same symmetric current J_S which is given by (29).

For each of these systems, we consider the large deviations of the time-averaged density, following Sect. 1.1.3. Large deviation principles of the form

$$\text{Prob}\left[\frac{1}{T}\int_0^T \rho_t^L(\cdot)\,dt \approx \rho(\cdot)\right] \asymp \mathrm{e}^{-TL^d I_2(\rho)} \tag{35}$$

apply in both reversible and irreversible models. This large deviation principle applies on taking the large-T limit after the hydrodynamic limit: one should take $L \to \infty$ before $T \to \infty$. To obtain bounds on I_2, we introduce the so-called level-2.5 large-deviation principle for the joint fluctuations of the empirical current and empirical measure [4,8]. That is,

$$\text{Prob}\left[\frac{1}{T}\int_0^T \rho_t^L(\cdot)\,dt \approx \rho(\cdot), \frac{1}{T}\int_0^T j_t^L(\cdot)\,dt \approx j(\cdot)\right] \approx \mathrm{e}^{-TL^d I_{2.5}(\rho,j)}. \tag{36}$$

If we assume that the paths that dominate the level-2.5 LDP are constant in time, the relevant rate function can be obtained from (24) as

$$I_{2.5}(\rho, j) = \frac{1}{4}\int_\Lambda (j - J(\rho))\cdot\chi(\rho)^{-1}(j - J(\rho))\,dx \tag{37}$$

if $\nabla\cdot j = 0$, and $I_{2.5} = \infty$ otherwise. The assumption of time-independent paths is equivalent to assuming that no dynamical phase transition takes place [6,11]. Using this assumption, we now calculate a bound (Proposition 3) for the level-2 rate functionals, which is analogous to (16) in the microscopic case.

Proposition 3 *Let the level-2.5 rate functional be given by* (37) *and let* I_2 *be the level-2 large deviation rate functional obtained from* $I_{2.5}$ *by contraction. We write* I_2^{rev} *for this rate functional if the current is symmetric,* $J = J_S$*, and we write* I_2^{irrev} *for the rate functional for the general case* $J = J_S + J_A$ *as in* (48)*. Then*

$$I_2^{\text{irrev}}(\rho) \geq I_2^{\text{rev}}(\rho). \tag{38}$$

Remark Note that this result will be strengthened later. We will obtain in equation (51) an exact identity for I_2^{irrev} as the sum of I_2^{rev} and a non-negative quantity.

Proof We write I_2 for I_2^{irrev}. The rate functional at level-2 can be obtained by a contraction of the level-2.5 rate functional,

$$I_2(\rho) = \inf_{j:\nabla\cdot j=0} I_{2.5}(\rho, j). \tag{39}$$

Note that $I_{2.5}(\rho, j)$ as given in Eq. (37) is [using (30)] equal to the sum of the following three summands:

$$\begin{aligned}&\frac{1}{4}\int_\Lambda (j - J_S(\rho))\cdot\chi(\rho)^{-1}(j - J_S(\rho))\,dx\\ &+\frac{1}{4}\int_\Lambda (j - J_A(\rho))\cdot\chi(\rho)^{-1}(j - J_A(\rho))\,dx - \frac{1}{4}\int_\Lambda j\cdot\chi(\rho)^{-1}j\,dx.\end{aligned} \tag{40}$$

The summand in the first line coincides with the symmetric rate functional $I_{2.5}^{\text{rev}}(\rho, j)$ and the second line is the part that corresponds to the anti-symmetric dynamics. Dropping the first summand in the second line (which is non-negative), we obtain

$$I_{2.5}(\rho, j) \geq \frac{1}{4}\int_\Lambda (j - J_S(\rho))\cdot\chi(\rho)^{-1}(j - J_S(\rho))\,dx - \frac{1}{4}\int_\Lambda j\cdot\chi(\rho)^{-1}j\,dx. \tag{41}$$

An expansion of the square shows that the right hand side is equal to

$$\frac{1}{4}\int_\Lambda J_S(\rho)\cdot\chi(\rho)^{-1}J_S(\rho)\,dx-\frac{1}{2}\int_\Lambda J_S(\rho)\cdot\chi(\rho)^{-1}j\,dx,$$

and the last summand vanishes under the assumption that $\nabla\cdot j=0$, as by Eq. (29)

$$\int_\Lambda J_S(\rho)\cdot\chi(\rho)^{-1}j\,dx=-\int_\Lambda\nabla\frac{\delta\mathcal{V}}{\delta\rho}\cdot j\,dx=\int_\Lambda\frac{\delta\mathcal{V}}{\delta\rho}\nabla\cdot j\,dx=0. \tag{42}$$

We obtain with (39) that

$$I_2(\rho)=\inf_{j:\nabla\cdot j=0}I_{2.5}(\rho,j)\geq\frac{1}{4}\int_\Lambda J_S(\rho)\cdot\chi(\rho)^{-1}J_S(\rho)\,dx.$$

To establish (39) we now show that the right hand side of this expression coincides with $I_2^{\mathrm{rev}}(\rho)$. Note that again for j such that $\nabla\cdot j=0$, by the same argument as in (42), the reversible level-2.5 rate functional is equal to

$$I_{2.5}^{\mathrm{rev}}(\rho,j)=\frac{1}{4}\int_\Lambda j\cdot\chi(\rho)^{-1}j\,dx+\frac{1}{4}\int_\Lambda J_S(\rho)\cdot\chi(\rho)^{-1}J_S(\rho)\,dx. \tag{43}$$

As one would expect for the reversible case, the infimum in (39) is clearly attained for a vanishing current ($j=0$), so that

$$I_2^{\mathrm{rev}}(\rho)=\frac{1}{4}\int_\Lambda J_S(\rho)\cdot\chi(\rho)^{-1}J_S(\rho)\,dx, \tag{44}$$

which completes the proof. □

Of course, given the acceleration at the microscopic scale, the result (38) that this acceleration is preserved at the hydrodynamic limit may not be surprising. However, we show below that the geometric structure underlying the MFT allows some stronger results for this acceleration to be established.

2.2.4 Splitting the Current

To understand the geometrical origin of (38) in more detail, we now show that as well as the decomposition (28), the antisymmetric current J_A has a further decomposition into two parts which are orthogonal to each other, and are both orthogonal to J_S. [Here, orthogonality should be understood in the sense of (30).]

We consider the problem

$$\nabla\cdot\big(\chi(\rho)\nabla\psi\big)=-\nabla\cdot J_A(\rho), \tag{45}$$

with the boundary condition $\psi=0$ on $\partial\Lambda$. For any fixed ρ (such that $\chi(\rho)$ and $J_A(\rho)$ are sufficiently regular) Eq. (45) has a unique strong solution ψ (see for example Theorem 6.24 in [20]). This solution ψ is therefore a functional of ρ we will denote with $\psi(\rho)$. Equation (45) motivates us to decompose $J_A(\rho)$ as

$$J_A(\rho)=-\chi(\rho)\nabla\psi(\rho)+J_F(\rho), \tag{46}$$

where $J_F(\rho)$ is a new vector field, which is again a functional of ρ. From (45) we see that

$$\nabla\cdot J_F(\rho)=0 \tag{47}$$

for all ρ.

We arrive at the following structure for the hydrodynamic current:

$$J(\rho) = J_S(\rho) - \chi(\rho)\nabla\psi(\rho) + J_F(\rho). \tag{48}$$

Of the three terms on the right hand side, the first is familiar as the symmetric current, while the third is divergence free and so does not transport any density. The remaining term (involving ψ) specifies how the density is transported by the antisymmetric current, and also determines the large deviations at level-2. The latter will be established below as a consequence of the following proposition.

Proposition 4 *The three terms on the right hand side of Eq.* (48) *are all orthogonal in the sense of Eq.* (30)*. Moreover, $J_S(\rho)$ and $-\chi(\rho)\nabla\psi(\rho)$ are orthogonal to all divergence free vector fields that vanish on the boundary.*

Proof Consider first the orthogonality between $J_F(\rho)$ and $\chi(\rho)\nabla\psi(\rho)$. One has $\psi(\rho)|_{\partial\Lambda} = 0$ so integration by parts yields

$$\int_\Lambda \chi(\rho)\nabla\psi(\rho)\cdot\chi^{-1}(\rho)J_F(\rho)\,dx = -\int_\Lambda \psi(\rho)\nabla\cdot J_F(\rho)\,dx = 0,$$

where the second equality follows from (47). Hence $J_F(\rho)$ and $\chi(\rho)\nabla\psi(\rho)$ are orthogonal in the sense of (30).

Following the same method but replacing ψ by $\delta\mathcal{V}/\delta\rho$ shows that $J_F(\rho)$ is orthogonal to $J_S(\rho) = -\chi(\rho)\nabla(\delta\mathcal{V}/\delta\rho)$, where we used $(\delta\mathcal{V}/\delta\rho)|_{\partial\Lambda} = 0$, as discussed in [9].

Finally, using the orthogonality relation (30) and $J_A(\rho) = -\chi(\rho)\nabla\psi(\rho) + J_F(\rho)$ yields

$$\int_\Lambda \chi(\rho)\nabla\psi(\rho)\cdot\chi^{-1}(\rho)J_S\,dx = \int_\Lambda J_F(\rho)\cdot\chi^{-1}(\rho)J_S(\rho)\,dx.$$

The right hand side vanishes by orthogonality of $J_S(\rho)$ and $J_F(\rho)$, so $\chi(\rho)\nabla\psi(\rho)$ is orthogonal to $J_S(\rho)$, as required. □

Combining Eq. (48) and Eq. (47), the dynamics of the density is given by

$$\partial_t\rho = \nabla\cdot\big(\chi(\rho)\big[\nabla\tfrac{\delta\mathcal{V}}{\delta\rho} + \nabla\psi(\rho)\big]\big). \tag{49}$$

The first term on the right hand side describes steepest descent (gradient flow) of the quasipotential, within a (modified) Wasserstein metric [1,27]. The second term describes a current that is orthogonal to the gradient flow (within the same metric), and leads to an evolution of ρ within the level sets of the quasipotential: this is the geometric result anticipated in Sect. 2.1.3, but in this hydrodynamic setting the geometrical objects are more explicit.

We now derive exact formulas for the level-2.5 and level-2 rate functionals based on the splitting in Proposition 4.

Proposition 5 *Let the level-2.5 large deviation rate functional be given by* (37)*. Further let ρ be such that Eq.* (45) *has a unique classic solution (up to a constant) and j such that $\nabla\cdot j = 0$. Then,*

$$\begin{aligned} I_{2.5}(\rho, j) =& \frac{1}{4}\int_\Lambda (j - J_F(\rho))\cdot\chi(\rho)^{-1}(j - J_F(\rho))\,dx \\ &+ \frac{1}{4}\int_\Lambda \nabla\tfrac{\delta\mathcal{V}}{\delta\rho}\cdot\chi(\rho)\nabla\tfrac{\delta\mathcal{V}}{\delta\rho}\,dx + \frac{1}{4}\int_\Lambda \nabla\psi(\rho)\cdot\chi(\rho)\nabla\psi(\rho)\,dx. \end{aligned} \tag{50}$$

Moreover, the level-2 rate functional is given by

$$I_2(\rho) = \frac{1}{4}\int_\Lambda \nabla\tfrac{\delta\mathcal{V}}{\delta\rho}\cdot\chi(\rho)\nabla\tfrac{\delta\mathcal{V}}{\delta\rho}\,dx + \frac{1}{4}\int_\Lambda \nabla\psi(\rho)\cdot\chi(\rho)\nabla\psi(\rho)\,dx. \tag{51}$$

Proof The proof of equation (50) follows from Proposition 4 and the representation of the rate functional (40). The second result (51) follows readily as $j = J_F(\rho)$ is the minimiser of (50). □

Note that these results are consistent with (43) and (44), where the minimising current was given by $j = 0$. In the general case, the minimising current is given by $j = J_F(\rho)$.

We moreover can recognise the first term on the right hand side of (51) as $I_2^{\text{rev}}(\rho)$, so the second term on the right hand side is an exact formula for the difference in rate for reversible and irreversible processes. This shows that the convergence rate for the irreversible process is strictly faster, unless the force $(-\nabla\psi)$ vanishes. We recognise this as a condition that the antisymmetric part of the current contributes to the time derivative of the density (otherwise the convergence to equilibrium of the density can not be accelerated).

Note that the objects $\nabla\frac{\delta\mathcal{V}}{\delta\rho}$ and $\nabla\psi$ should be interpreted as forces acting in the space of densities. In order to sustain a large deviation of the density, the stochastic forces within the system must act to resist these (deterministic) forces. One sees from (51) that the probability of this rare event (or large deviation) is given by the norms of the two forces, within a metric that depends on the mobility χ.

2.2.5 An Example

We have discussed the status of the MFT as a theory for the hydrodynamic limit of interacting particle systems. For a concrete example of this approach, we consider an interacting particle model known as the zero-range process (ZRP) [38]. A microscopic description of the ZRP is given in Sect. 3.1. For the purposes of this section, the important features of the ZRP are that its hydrodynamic limit is described by the MFT and that irreversible ZRPs have local quasipotentials of the form (32). This latter fact allows straightforward comparison between reversible and irreversible models with the same quasipotential.

The hydrodynamic limit of the ZRP is a non-linear drift-diffusion

$$\partial_t\rho = \Delta\phi(\rho) - \nabla\cdot\big(\phi(\rho)E\big), \tag{52}$$

where ϕ is a function that depends on the local density [that is, $\phi(\rho)(x) = \phi(\rho(x))$], and E is a drift term. The specific function ϕ that appears in the MFT depends on how the particles interact within the ZRP. A formal derivation of this hydrodynamic limit can e.g. be found in [9]. If $\phi(\rho) = \rho$, then the model corresponds to drift-diffusion of non-interacting particles.

One sees immediately from (52) that the hydrodynamic current is given by (26) with $\chi(\rho) = \phi(\rho)I$ and $D(\rho) = \phi'(\rho)I$, where I is the identity matrix. Moreover, the quasipotential for the ZRP is given by (32) with $f'(\rho) = \log\phi(\rho)$, consistent with (34). The ZRP may be either reversible or irreversible: one sees that reversible ZRPs lead to $E = -\nabla V$ for some potential V. In this case (34) shows that $V(x) = \log(\phi(\overline{\rho}(x))) + \lambda$, where $\overline{\rho}$ is the steady state density profile and λ is a constant (independent of x). Hence one identifies the irreversible current as $J_A(\rho) = J(\rho) - J_S(\rho) = \phi(\rho)\big[E + \nabla\log\phi(\overline{\rho})\big]$.

Examining the rate function (51) for the specific case of the ZRP, one can interpret the result as a generalisation of a result in [35]. One has $\delta\mathcal{V}/\delta\rho = \log\phi(\rho) - \log\phi(\overline{\rho})$. Hence

$$I_2(\rho) = \int\Big(\big|\nabla\log\big(\tfrac{\phi(\rho)}{\phi(\overline{\rho})}\big)\big|^2 + |\nabla\psi(\rho)|^2\Big)\phi(\rho)\,dx, \tag{53}$$

where ψ is the solution of $\nabla\cdot\big(\phi(\rho)\nabla\psi\big) = -\nabla\cdot[\phi(\rho)(E + \nabla\log\phi(\overline{\rho}))]$. If we now consider the special case $\phi(\rho) = \rho$ then we recover the same rate function as in Theorem 2.2 of Ref. [35]: the non-gradient force C in that work is here replaced by $E + \nabla\log\overline{\rho}$ (note that

this is independent of ρ). The condition that $\nabla \cdot (\overline{\rho} C) = 0$—which ensures that the invariant measure is unchanged by breaking detailed balance—is satisfied within the MFT because $\nabla \cdot J_A(\overline{\rho}) = 0$ and setting $\phi(\rho) = \rho$ yields $J_A(\overline{\rho}) = \overline{\rho}(E + \nabla \log \overline{\rho})$.

Note however the setting discussed in this work is different to that in [35]: here we consider the hydrodynamic limit of many particles on a lattice while that work considers a single particle in a compact manifold without boundary. For non-interacting particles, the result is the same: the reason that for the many-particle system, the rate function I^N associated with all the particles undergoing the same rare fluctuation is equal to NI^1. So the only difference between the one-particle and many-particle systems arises in the prefactors (speeds) of the large deviation principles (35), (36).

3 Application to the Zero-Range Process, and Numerical Results

3.1 The Zero-Range Process

The ZRP [38] is a system in which interacting particles move on a finite lattice $\Lambda_L = \{0, \ldots, L-1\}^d \subseteq \mathbb{Z}^d$ where $L \in \mathbb{N}$ is the linear system size. The particles are assumed to be indistinguishable and each particle is located at one of the sites $x \in \Lambda_L$. The number of particles on site x is $\eta(x)$ and the configurations of the system are $\eta = (\eta(x))_{x \in \Lambda_L}$. We will assume that the total number of particles is conserved such that no particles are added or removed over time.

The interaction of the particles is encoded in a function $g(k)$, with $g(0) = 0$. The rate of particle transfer from site x to site y is $g(\eta(x))c(x \to y)$, where the function c determines the connectivity of the sites. The case $g(k) = k$ corresponds to non-interacting particles. The model is referred to as zero-range because particles interact only when they are on the same site. For example, if $g(k) = k^\alpha$ for $k > 0$, then $\alpha < 1$ means particles on the same site attract each other (suppressing jumps away from that site) while $\alpha > 1$ means that particles on the same site tend to repel each other.

3.1.1 Reversible and Irreversible ZRP

The behaviour of the ZRP depends strongly on the choice of the connectivity function c as well as the interaction function g. We assume that particles hop only to nearest neighbour sites, so $c(x \to y) > 0$ only if x and y are nearest neighbours. At the boundaries of the lattice, the system has either reflecting boundaries (particles cannot leave the lattice) or periodic boundaries.

It is easily verified that the model obeys the detailed balance condition (6) if one takes (for nearest neighbour sites)

$$c(x \to y) = \mathrm{e}^{\frac{1}{2}[V(x) - V(y)]} \tag{54}$$

for some potential function V. In this case the model is reversible.

To arrive at a class of irreversible models, we take

$$c(x \to y) = \mathrm{e}^{\frac{1}{2}[V(x) - V(y)]} + k_{x,y} \mathrm{e}^{V(x)} \tag{55}$$

with $k_{x,y} = -k_{y,x}$. In this case positivity of transition rates requires $|k_{x,y}| < \mathrm{e}^{-\frac{1}{2}[V(x)+V(y)]}$ for all x, y. We show below that taking $k \neq 0$ corresponds to breaking of detailed balance, in the sense of (7).

3.1.2 Generator and Invariant Measure

We denote the configuration of the ZRP at time t with η_t. The generator acts on the test function f as

$$\mathcal{L}f(\eta) = \sum_{x,y\in\Lambda_L} (f(\eta^{x,y}) - f(\eta))g(\eta(x))c(x\to y). \tag{56}$$

Here $\eta^{x,y}$ denotes the configuration obtained from η by removing one particle from position x and adding it at position y. If $\eta(x) = 0$ we simply set $\eta^{x,y} = \eta$ and hence leave the configuration unchanged.

Note that the ZRP as defined so far is reducible, since the number of particles is a conserved quantity under the dynamics. This setting is useful because it is easily verified (directly from the definition (56) and using that the invariant measure π satisfies $\sum_\eta \pi(\eta)\mathcal{L}f(\eta) = 0$ for all f) that the reversible model with rates defined in (54) has a family of invariant measures, the so called grand-canonical measures, which are parameterised by the chemical potential λ and given by

$$\pi^{\varphi}_{\text{grand}}(\eta) = \prod_{x\in\Lambda_L} \frac{\varphi(x)^{\eta(x)}}{z(\varphi(x))g!(\eta(x))} \tag{57}$$

with the fugacity $\varphi(x) = \mathrm{e}^{-V(x)-\lambda}$ for some $\lambda \in \mathbb{R}$; the notation $g!(k)$ indicates the generalised factorial $g!(k) := \prod_{i=1}^{k} g(i)$ [with $g!(0) = 1$] and $z(\varphi) = \sum_{k=0}^{\infty} \frac{\varphi^k}{g!(k)}$ is a normalisation constant [19,28]. We here assume that V, λ and $g(\cdot)$ are such that $z(\varphi(x)) < \infty$ for all $x \in \Lambda_L$. This is in particular the case for any V and λ, when $g(\cdot)$ satisfies $g(k) \geq ck$ for some constant $c > 0$ [28].

On restricting the model to a fixed number of particles N, the invariant measure π (which is called the canonical measure) can be obtained by a conditioning of (57). Note that (57) has the structure of a product measure. Also if $g(k) = k$ then one recovers the case of non-interacting particles and the local marginals of (57) are Poisson distributions.

To make the comparison between reversible and irreversible models described in Sect. 2.1, we require an irreversible model whose invariant measure is (57). Again using that $\sum_\eta \pi(\eta)\mathcal{L}f(\eta) = 0$ for all f, we take $f = \eta(x)$ to be the number of particles on site x, from which we see that the irreversible rates (55) are also consistent with the invariant measure (57) if we take

$$\sum_{y:y\sim x} (k_{x,y} - k_{y,x}) = 0 \quad \text{for all } x, \tag{58}$$

where the notation $y \sim x$ indicates that sites x and y are nearest neighbours. (If we imagine a system with just one particle, this constraint states that the rate of hopping onto site x is balanced by the rate of hopping away from that site. For the ZRP, this same balance condition ensures that the invariant measure (57) is still valid even for many interacting particles).

Finally then, the conditions on the perturbations $k_{x,y}$ required for a meaningful comparison between reversible and irreversible models can be summarised as:

$$\sum_{y:y\sim x} k_{x,y} = 0, \quad k_{x,y} = -k_{y,x}, \quad \text{and } |k_{x,y}| < \mathrm{e}^{-(V(x)+V(y))/2}. \tag{59}$$

The rates $k_{x,y}$ can be interpreted as elements of a matrix, which coincides (up to the factor $1/2$) with the vorticity matrix Γ introduced in [10].

In terms of the splitting (7) the symmetric part of the dynamics is given by $c_s(x \to y) = \mathrm{e}^{\frac{1}{2}[V(x)-V(y)]}$ and the anti-symmetric part by $c_a(x \to y) = k_{x,y}\mathrm{e}^{V(x)}$, such that the symmetric part (corresponding to $\mathcal{L}_S$) is independent of $k_{x,y}$.

3.1.3 Hydrodynamic Limit

The hydrodynamic limit of the ZRP is defined as follows. For a ZRP on a lattice Λ_L with L^d sites, one takes $N = \lfloor \rho_0 L^d \rfloor$ particles, where ρ_0 is the average density. The lattice Λ_L is mapped into the domain $[0, 1]^d$ by identifying each site $x \in \Lambda_L$ with a position $\tilde{x} \in \Lambda$ with $\Lambda = [0, 1]^d$. Hence the site x with integer co-ordinates $(i, j, \dots)$ has a position $\tilde{x} = (i/L, j/L, \dots)$. Roughly speaking, the density $\rho_t(\tilde{x})$ in the MFT is equal to the typical number of particles on site x, and the normalisation of the density is $\int_\Lambda \rho_t(\tilde{x})\, d\tilde{x} = \rho_0$. The hydrodynamic limit corresponds to a sequence of models in which $L \to \infty$ at fixed ρ_0, so $N \to \infty$.

The hydrodynamic limit corresponds to observing a system on increasingly large length and time scales. Note that since the number of sites in Λ_L is diverging (proportional to L^d) in the hydrodynamic limit, the diffusion constant for a single particle (in Λ) vanishes as L^{-2}. For this reason, when the lattice Λ_L is mapped into the fixed domain Λ, it is also convenient to scale the hop rates for all particles, by taking $c(x \to y) \to L^2 c(x \to y)$. This ensures that the diffusive behaviour characteristic of the hydrodynamic limit is observed, and the hydrodynamic limit is consistent with MFT.

To fix the hop rates between sites in the ZRP, one fixes a smooth potential function $\tilde{V} : \Lambda \to \mathbb{R}$ on the hydrodynamic scale, and one considers a sequence of ZRPs of increasing sizes L with potential functions $V(x) = \tilde{V}(\tilde{x})$, where $\tilde{x}$ is the image in Λ of the discrete site $x \in \Lambda_L$. Similarly one fixes a vector field $\tilde{k} : \Lambda \to \mathbb{R}^d$ and takes $k_{x,y} = \tilde{k}(\tilde{x}) \cdot (\tilde{y} - \tilde{x})$ where the dot indicates a scalar product in $\mathbb{R}^d$.

The relation between the ZRP and the MFT is discussed in e.g. [7], [22] and in the review paper [9]. In particular, for both reversible and irreversible ZRPs one arrives at the situation described in Sect. 2.2.5. The hydrodynamic limit (52) depends on the drift function $E : \Lambda \to \mathbb{R}^d$ which is given by $E(\tilde{x}) = -\nabla \tilde{V}(\tilde{x}) + \tilde{k}(\tilde{x})$.

The MFT description of the ZRP also depends on a function ϕ which can be obtained as the solution of

$$\rho = \sum_{k=1}^{\infty} \frac{k\,\phi(\rho)^k}{z(\phi(\rho))g!(k)}. \tag{60}$$

We identify the right hand side of this equation as the mean local density associated with the measure (57), at fugacity $\varphi = \phi(\rho)$.

The quasipotential $\mathcal{V}$ for the ZRP is given by [9],

$$\mathcal{V}(\rho) = \int_\Lambda \left[\rho(x) \log\left(\frac{\phi(\rho(x))}{\phi(\bar{\rho}(x))} \right) - \log\left(\frac{z(\phi(\rho(x)))}{z(\phi(\bar{\rho}(x)))} \right) \right] dx. \tag{61}$$

3.2 Simulation Results

We present numerical results for one-dimensional and two-dimensional systems, showing how breaking detailed balance [that is, taking $k_{x,y} \neq 0$ in (55)] accelerates convergence to equilibrium. The simulations are performed using the Gillespie algorithm [21]. The results illustrate several aspects of the theoretical analysis in Sect. 2. First, the results of that section

do not rely on how detailed balance is broken: we show that there are several possible choices and discuss their consequences. Second, our numerical results show in what contexts we expect to see significant acceleration of the dynamics on breaking detailed balance, and in what contexts we expect the acceleration to be mild.

In all cases, we show results that are scaled to be consistent with the hydrodynamic limit. That is, we map the lattice Λ_L into $[0, 1]^d$ and we rescale the microscopic hop rates by a factor of L^2 so as to recover diffusive behaviour in the hydrodynamic limit.

In practical situations where the rate of convergence to equilibrium is important, a common situation is that the potential function V is not convex, but includes several (or many) minima, separated by high barriers. From a physical perspective, the temperature of our systems is a parameter that has been absorbed into the function V. In general, high barriers are linked with long (Arrhenius) time scales that are proportional to $\mathrm{e}^{\Delta V}$. In order to understand whether breaking detailed balance can accelerate convergence in such non-convex problems, we consider cases where the function V has two minima, with longest time scale in the system corresponding to motion between these minima.

3.2.1 Characterisation of Convergence

We perform numerical simulations starting from a fixed (deterministic) initial condition η_0. To analyse convergence to equilibrium, we perform numerical simulations of the ZRP, and we track the time-dependence of several different quantities. For any configuration η, the mean potential energy is

$$\langle \eta, V \rangle = \sum_{x \in \Lambda_L} \eta(x) V(x). \tag{62}$$

We generate several trajectories (sample paths) η_t of the ZRP and we estimate the mean potential energy

$$\hat{V}(t) = \mathbb{E}_{\mu_0}(\langle \eta_t, V \rangle) \tag{63}$$

by taking the mean value of $\langle \eta_t, V \rangle$ over these trajectories. For systems of non-interacting particles (where $\phi(\rho) = \rho$), we also estimate the macroscopic relative entropy as

$$D(t) = \sum_{x \in \Lambda_L} \mathbb{E}_{\mu_0}(\eta_t(x)) \log\left(\frac{\mathbb{E}_{\mu_0}(\eta_t(x))}{\mathbb{E}_{\pi}(\eta(x))} \right), \tag{64}$$

which can be seen as an approximation to the quasipotential, which is for an independent random walk given by

$$\mathcal{V}(\rho_t) = \int_\Lambda \left[\rho_t(x) \log\left(\frac{\rho_t(x)}{\bar{\rho}(x)} \right) + \rho_t(x) - \bar{\rho}(x) \right] dx = \int_\Lambda \rho_t(x) \log\left(\frac{\rho_t(x)}{\bar{\rho}(x)} \right) dx,$$

where we used the fact that $z(\varphi) = \mathrm{e}^{-\varphi}$ in (61) and the last identity follows from the fact that the density is conserved: $\int_\Lambda \rho_t(x) dx = \int_\Lambda \bar{\rho}(x) dx$.

For numerical purposes, we estimate $\mathbb{E}_{\mu_0}(\eta_t(x))$ as the average occupancy of site x over the sample paths that we generate, and we calculate $\mathbb{E}_{\pi}(\eta(x))$ by direct construction of the invariant measure (whenever possible). Finally, we estimate the Gibbs entropy

$$S(t) = -\sum_x \mathbb{E}_{\mu_0}(\eta_t(x)) \log \mathbb{E}_{\mu_0}(\eta_t(x)), \tag{65}$$

which is large if particles are delocalised throughout the system, and small if they are concentrated on a small number of sites. Again, we estimate $\mathbb{E}_{\mu_0}(\eta_t(x))$ as the average occupancy of site x over the sample paths that we generate, which provides an estimator of S.

These three quantities $\hat{V}$, D, S all converge as a function of time to stationary values, providing differing information as to the rates of convergence. Note that for non-interacting particles, $\overline{\rho}(x) = \mathbb{E}_\pi(\eta(x)) = \mathrm{e}^{-V(x)}/z$ for some constant z, so $D(t) = -S(t) + \hat{V}(t) + \log z$.

3.2.2 One-Dimensional Case: Results

We consider periodic boundaries for a model on a one-dimensional strip, this is equivalent to motion on the perimeter of a circle (flat torus in one dimension). In this case condition (59) requires $k_{x,x+1} = k_{x-1,x}$, so we set $k_{x,x+1} = c$ with some constant c that is independent of x. The choice $c > 0$ corresponds to a fixed force $c\,\mathrm{e}^V$ that is forcing the particles to travel around the circle. For a hydrodynamic limit consistent with macroscopic fluctuation theory, we require c to vary with the system size L as $c = E/L$ with E a fixed constant [9].

We note in passing that the use of periodic boundaries is essential for breaking balance in these closed systems: on a finite strip with reflecting boundary conditions, (59) has no solutions except $k_{x,y} = 0$ so there is no way to break detailed balance.

Thus, returning to the case with the periodic boundaries, the generator is

$$\begin{aligned}\mathcal{L}f(\eta) = \sum_{x=0}^{L-1}\Big[&\big(f(\eta^{x,x+1}) - f(\eta)\big)\,L^2 g(\eta(x))\big(\mathrm{e}^{(V(x)-V(x+1))/2} + (E/L)\mathrm{e}^{V(x)}\big)\\ &+ \big(f(\eta^{x,x-1}) - f(\eta)\big)\,L^2 g(\eta(x))\big(\mathrm{e}^{(V(x)-V(x-1))/2} - (E/L)\mathrm{e}^{V(x)}\big)\Big],\end{aligned} \tag{66}$$

where the addition is periodically extended on $\Lambda_L = \{0, \ldots, L-1\}$, i.e., $(L-1)+1 = 0$ and $0 - 1 = L - 1$. We take $g(k) = k$ so that the particles do not interact. The potential is

$$V(x) = A\sin(4\pi x/L) - B\cos(2\pi x/L) \tag{67}$$

with $A = 3/2$ and $B = 3/4$ so that the global minimum of the potential is at $\hat{x} \approx 0.888$ with $V \approx -2.052$. The height of the barrier is approx 2.609. The initial condition has all particles on a single site, $x_0 = L/4$, in the vicinity of the secondary minimum. The stationary state has $\overline{\rho}(x) = \mathbb{E}_\pi(\eta(x)) \propto \mathrm{e}^{-V(x)}$ with a proportionality constant determined by the total density (which in this case is $z \approx 2\,377$). The parameter E in Eq. (66) is set to $E = 36$. For the lattice size $L = 300$, the maximal value allowed for E to ensure that $c_s + c_a \geq 0$ is slightly above 38.4. In principle one can choose larger values for E by increasing the lattice size L.

The results in Fig. 1 are for a domain of size $L = 300$; we also compared this to simulations for $L = 150$, $L = 300$ and $L = 450$ for the value $E = 18$ (to ensure positiveness of the transition rates for $L = 150$). We found the results to be qualitatively very similar, see the bottom right panel in Fig. 1. Figure 1 shows the convergence to equilibrium of the mean potential energy and the entropy. One sees that convergence of both the energy and the entropy is significantly faster when detailed balance is broken. To illustrate the mechanism for this effect, Fig. 2 shows how the mean density $\mathbb{E}_{\mu_0}(\eta_t(x))$ varies with time. In the irreversible case, the non-gradient part of the drift force E acts to the right and is equal to $c\,\mathrm{e}^V$, so it is large near the maxima of the potential. This prevents the system from becoming localised in the secondary (local) minimum and aids convergence to the steady state. By contrast, in the reversible system, the particles need to *diffuse* over the maxima of the potential, which is a slower process. This difference explains the much faster convergence

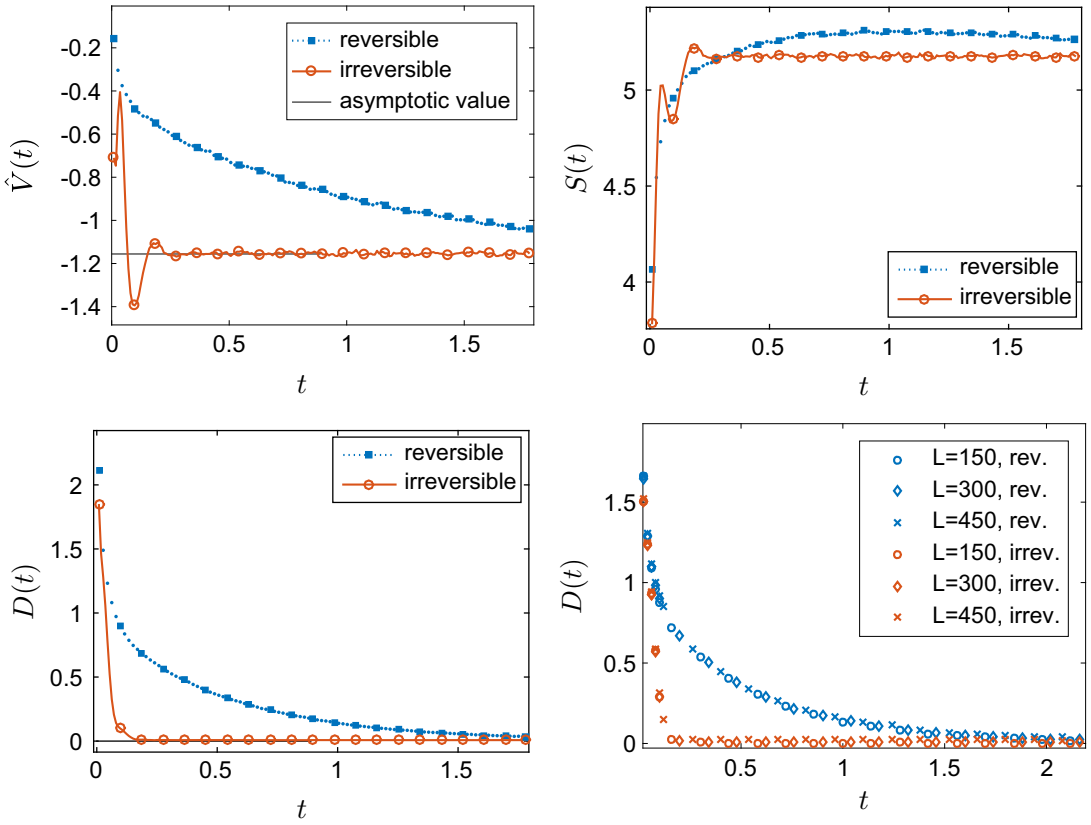

Fig. 1 One-dimensional simulation for independent particles on a *circle* with $L = 300$ sites, comparing reversible and irreversible drift-diffusion processes as described in the main text with the potential (67). *Top row and bottom left* plot of the test observables average energy $\hat{V}$, Gibbs entropy S and relative entropy D for $E = 36$. *Bottom right* plot of the relative entropy D for different system sizes $L = 150, 300, 450$, all for $E = 18$. As predicted by the hydrodynamic equation, varying the system size at fixed E and rescaling time by a factor of L^2 leads to limiting behaviour independent of L. All results were obtained by averaging over 20, 000 individual particle trajectories

to the steady state observed in Fig. 1. The overshoot of the entropy for the reversible case in Fig. 1 occurs because the state where the particles are distributed evenly between the two minima has a higher entropy S than the steady state (where they are localised primarily in the global minimum). The state where the particles are distributed evenly between the minima is an example of a situation where the gradient of the free energy is small (within the relevant metric), so that steepest descent of the free energy leads to slow changes in the density.

Note also that (64) implies that $D(t) \to 0$ at long times, as the system converges to its steady state. However, in Fig. 1 one sees that our estimate of $D(t)$ converges instead to a small positive constant. This offset arises because our estimator of $D(t)$ is biased: it is based on m independent numerical simulations (each with N particles) and the expectation value of our estimator converges to $D(t)$ only as $m \to \infty$. Specifically, we estimate $\mathbb{E}_{\mu_0}(\eta_t(x))$ as $\vartheta_t(x) = m^{-1} \sum_{k=1}^{m} \eta_t^k(x)$ where $\eta_t^k(x)$ is the number of particles on site x at time t in the k-th simulation. Inserting this estimate into the (nonlinear) expression (64), it is easily shown that the resulting estimator of $D(t)$ has in general a finite bias. However, as $m \to \infty$, ϑ obeys a law of large numbers and converges almost surely to $\mathbb{E}_{\mu_0}(\eta_t(x))$—hence our estimator converges to $D(t)$ as $m \to \infty$.

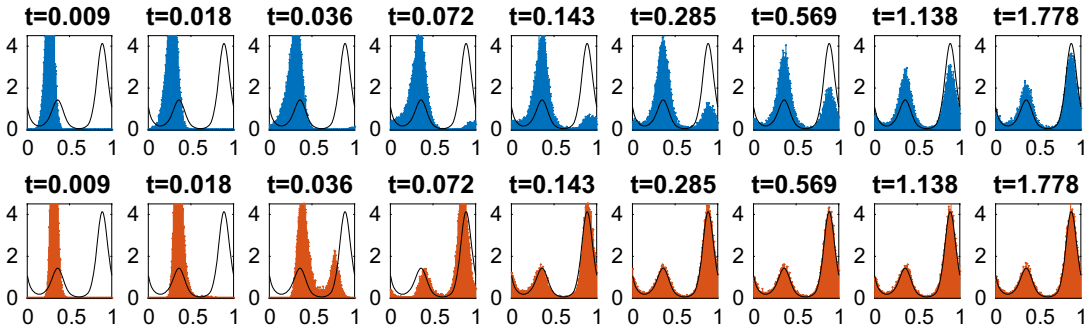

Fig. 2 One-dimensional simulation for independent random walk on a *circle* with the potential (67). Configuration at different times for the reversible (*top row*) and the irreversible (*bottom row*) process with drift 'to the right' and steady state (*in black*). x-axis: position. y-axis: averaged number of particles. In the irreversible case, $E = 36$

3.2.3 One-Dimensional Case: Discussion

This one-dimensional model is useful for illustrative purposes and establishes the general principles derived in Sect. 2. However, the restriction to one dimension means that detailed balance can only be broken by applying a driving force $c\,\mathrm{e}^V$ (otherwise the invariant measure would be changed). If barriers are large, one sees that the driving force near the top of the barrier must be very large indeed: it is hard to see how this can be realised in practical applications. Physically, the idea is to drive a constant current around the periodic system, and this requires the drift velocities (and hence forces) to be largest at the top of any barriers, where the density is least. In this sense, it is perhaps not surprising that by applying large forces to quickly drive particles over all barriers in the system, one can significantly speed up mixing of the particles between the two minima of the potential.

For these reasons, we turn to a two-dimensional system, where there are many more ways of breaking detailed balance while preserving the same invariant measure.

3.2.4 Two Dimensional Case: Model and Results

In two dimensions, there is considerably more freedom in the choice of the rates $k_{x,y}$. If one again assumes periodic boundaries, it is always possible to have all non-gradient forces acting in a single direction: for example $k_{x,x+e_1} = c$ where e_1 is a lattice vector, as in the previous one-dimensional example. However, this requires driving forces that depend exponentially on the value of the potential, as in one dimension. We therefore pursue a different strategy.

Denoting the Euclidean basis for Λ_L with e_1, e_2, Eq. (59) implies that both $k_{x,x\pm e_j} = -k_{x\pm e_j,x}$ and $k_{x,x+e_1} + k_{x,x-e_2} + k_{x,x-e_1} + k_{x,x+e_2} = 0$ have to be satisfied. One way to choose appropriate $k_{x,y}$ is to consider the plaquettes of the square lattice as in Fig. 3 and to define a *vorticity* W at the centre of each plaquette. The value of W on the plaquette centred at $x + \frac{1}{2}(e_1 + e_2)$ is $W(x)$. One then can choose the rates $k_{x,y}$ as the following differences:

$$\begin{aligned} k_{x,x+e_1} &= W(x-e_2) - W(x) \\ k_{x,x-e_2} &= W(x-e_1-e_2) - W(x-e_2) \\ k_{x,x-e_1} &= W(x-e_1) - W(x-e_1-e_2) \\ k_{x,x+e_2} &= W(x) - W(x-e_1) \end{aligned} \tag{68}$$

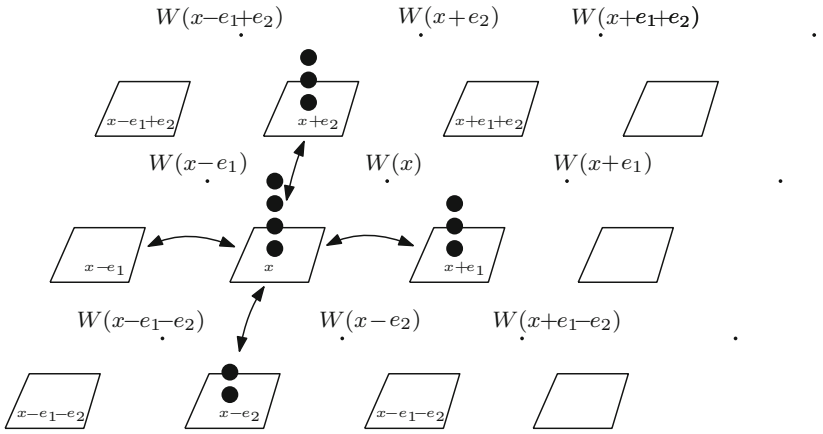

Fig. 3 The function W defined on the plaquettes; e_1 and e_2 are the Euclidean basis vectors

This choice satisfies both conditions $k_{x,y} = -k_{y,x}$ and $\sum_y k_{x,y} = 0$. The quantity W can be identified as a vorticity, in the sense that taking $W(x) = W_0\delta_{x_0,x}$ with $W_0 > 0$ causes particles to circulate clockwise around plaquette x_0.

Any choice of the function W is possible, and should lead to acceleration of the dynamics, following the theoretical analysis of Sect. 2. Here we concentrate on a case where W is related to the potential V, so that the rates $c(x \to y)$ depend only on the gradients of the potential in the vicinity of site x. (The physical idea is that particle motion is naturally sensitive to local potential gradients since these correspond to forces acting on the particles. On the other hand, the motion of a particular particle should not be sensitive to the total energy V, since this depends on the state of the system far away from that particle). To arrive at forces that depend only on potential gradients, we take $W(x) = a \cdot \exp(\frac{1}{4}[V(x) + V(x + e_1) + V(x + e_2) + V(x + e_1 + e_2)])$, where a is a parameter that sets the scale of the vorticity.

On taking the hydrodynamic limit, this gives rise to the driving force

$$E(\tilde{x}) = -\nabla\tilde{V}(\tilde{x}) + a[e_1\nabla_2\tilde{V}(\tilde{x}) - e_2\nabla_1\tilde{V}(\tilde{x})], \tag{69}$$

where $a > 0$ (recall from Sect. 3.1.3 that $\tilde{x}$ is the image in Λ of the discrete site $x \in \Lambda_L$). We recognise the second term on the right hand side as a force that is obtained by rotating ∇V clockwise by $\pi/2$ radians, so that it acts to drive the system around the level sets of V.

The following simulations are on a two dimensional closed domain with $L = 140$ and zero flux at the boundary, i.e., the domain has $140 \times 140 = 19\,600$ sites and the particles cannot leave the domain.

We consider three different ZRPs, corresponding to different choices for $g(k)$. Firstly, we consider the linear case (independent particles), where $g(k) = k$. We further consider the superlinear case with $g(k) = k^{3/2}$, such that the particles repel each other (the hop rates away from site x is increased when that site contains more particles). Finally we investigate the sublinear case with $g(k) = k^{5/6}$ in which the particles prefer to cluster together. For each setting, we simulated the process with both reversible and irreversible dynamics, with $L^2/2 = 9\,800$ particles averaged over 16 simulations. The potential, which is also depicted in Fig. 4, is for shifted coordinates $x = (x_1, x_2) \in [-1/2, 1/2]^2$ given by

$$V(x_1, x_2) = A(x_1^2 - B)^2 + Cx_2^2 + Dx_1 \tag{70}$$

with a cut-off at a given height V^*. For the simulations we chose the parameters $A = 500$, $B = 0.085$, $C = 30$, $D = 2.5$ and $V^* = 5$ (that is, the potential used is $\max(V(x_1, x_2), V^*)$). The parameter in (69), which sets the strength of the non-gradient term of the driving force,

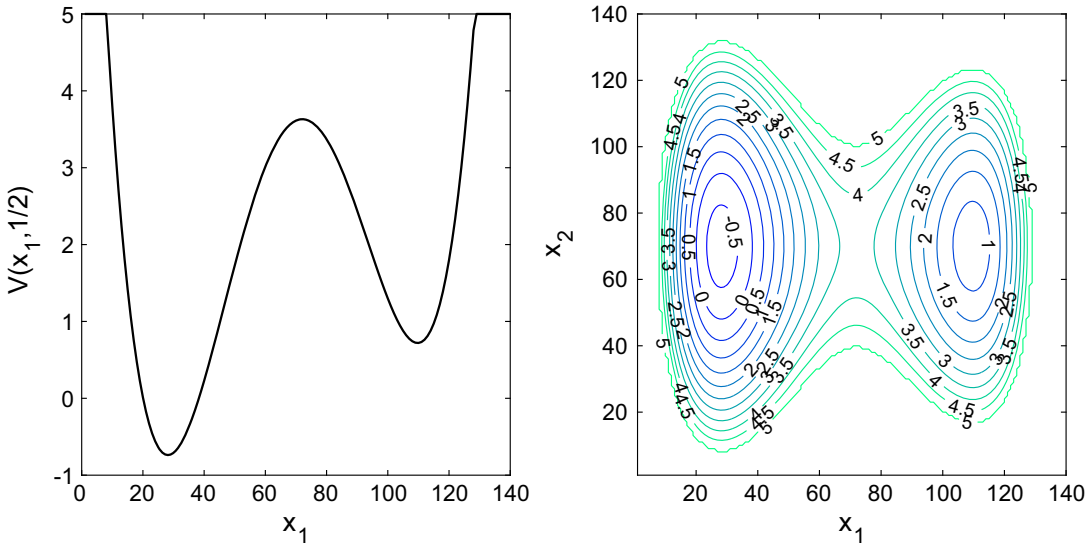

Fig. 4 *Left* x_1-cross-section of $V(x)$ as given in (70) for $x_2 = 1/2$. *Right* level sets of $V(x)$

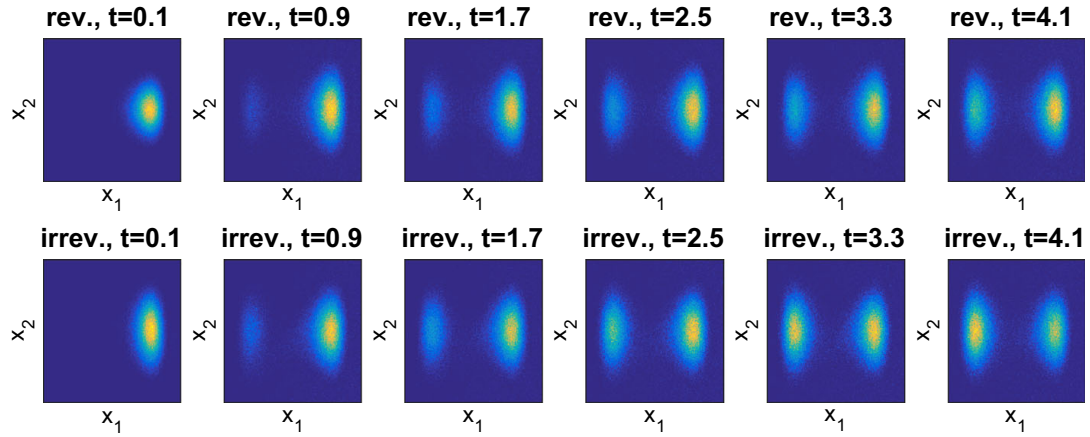

Fig. 5 Configuration for $g(k) = k$ with the potential (70) at different times. (*Dark*) *blue* means low number of particles, *yellow* means many particles. *Top row* reversible process. *Bottom row* irreversible process (Color figure online)

was set to $a = 0.4$. This value is again close to the maximal allowed value (which is slightly above 0.405).

For all simulations, the particles start at position $(0.5, 0.75) \in [0, 1]^2$ close to the local minimum of the double well potential. The particles then try to leave this well and move to the global minimum (on the left) as can be seen in the plots in Fig. 5 for the linear case. The test observables for the linear/superlinear/sublinear case can be found in Figs. 6, 7 and 8, respectively. Depending on the chosen configuration, the simulation time on a HPC node with 16 cores using Matlab took between 10 and 13.5 hours.

As in the one-dimensional case, the particles are under the irreversible dynamics able to leave the minimum faster than it is the case for reversible dynamics (compare the bottom row with the top row in Fig. 5).

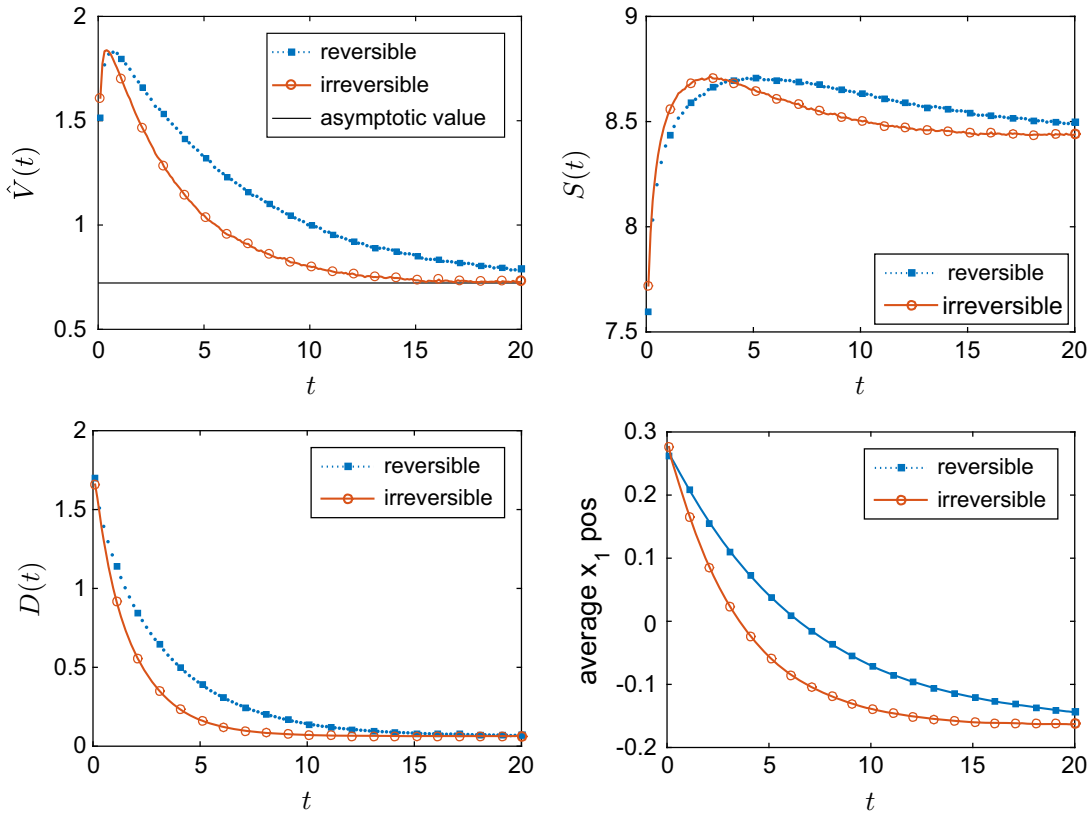

Fig. 6 ZRP with $g(k) = k$ (independent random walk) with initial position of all particles in the local minimum. Average energy $\hat{V}$, the Gibbs entropy S, the relative entropy D and the average x_1-position of particles. The initial position of the particle is at a fixed position in the local (but not global) minimum of the potential (70). The domain size is $L^2 = 140^2$ and we averaged over 16 simulations consisting of 9800 particles each

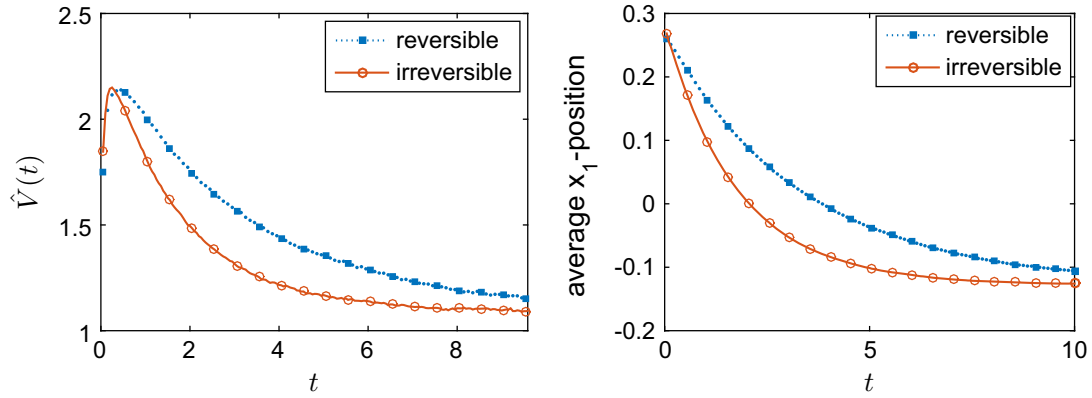

Fig. 7 ZRP with $g(k) = k^{3/2}$ and the particles are started in the local minimum. *Left* average energy $\hat{V}$. *Right* average x_1-position

3.2.5 Two Dimensional Case: Discussion

We close this section with Table 1, which quantifies the acceleration in the models where particles attract, repel, or have no interactions. For this, we consider the average energy $\hat{V}$ and the average x_1 position of the particles. Assuming that the final values of these observables in the irreversible simulations are close to their steady-state values, we consider the distance

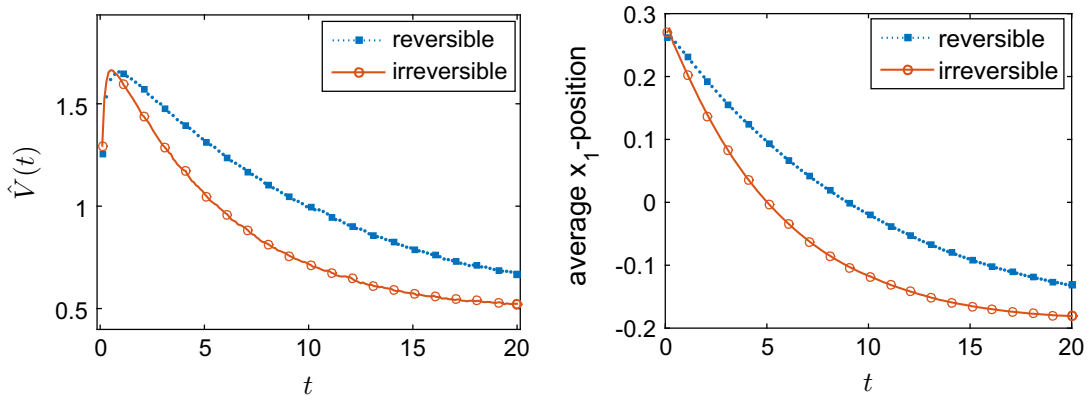

Fig. 8 ZRP with $g(k) = k^{5/6}$ and the particles are started in the local minimum. We again plot the average energy $\hat{V}$ and the average x_1-position

Table 1 Table of the absolute times t_s for the reversible process (left) and ratios between times of the reversible and irreversible process t_s/t_a (right) to reach the distances $\Delta\hat{V} = 0.3$ and $\Delta x_1 = 0.2$, respectively

t_s	$\Delta\hat{V}$	Δx_1	t_s/t_a	$\Delta\hat{V}$	Δx_1
$g(k) = k$	2.38	1.30	$g(k) = k$	1.83	1.79
$g(k) = k^{3/2}$	1.11	0.58	$g(k) = k^{3/2}$	1.78	1.86
$g(k) = k^{5/6}$	3.55	2.03	$g(k) = k^{5/6}$	1.77	1.80

$\Delta\hat{V}$ (resp. Δx_1) of both the reversible and irreversible process and keep track of the first time where the distance is below a given threshold. Denoting this time for the reversible process with t_s and for the irreversible process with t_a, we can use the ratio t_s/t_a as an estimator for the acceleration.

From the data in the table, on sees that the processes are typically accelerated by factors about 1.75 independent of the choice of $g(k)$. We checked different thresholds (here we displayed $\Delta\hat{V} = 0.3$ and $\Delta x_1 = 0.2$) which all lead to the same conclusions.

These are significant accelerations, although considerably less than the dramatic speedup of order 10 observed in one dimension. However, the physical mechanisms for the acceleration are different in the two cases. In one dimension, the drift forces which act to push particles up and over the barrier, so the forces are very large at the top of the barrier. In two dimensions, the effect is more subtle: returning to Fig. 4 and recalling that the drift force in (69) is obtained by a rotation of the potential gradient, one sees that in the vicinity of the saddle point of the potential, there is a net drift to the left in the top part of Fig. 4b, and a drift to the right in the bottom part. A natural analogy is a gentle stirring motion that happens in the vicinity of the saddle point, and tends to accelerate mixing. This seems a much more plausible mechanism for accelerating convergence to equilibrium in practical situations, compared with the large forces required in one dimension.

Finally, we note that transport between the minima of a non-convex potential energy always involves a slow time scale proportional to $e^{\Delta V}$, since a particle must still reach the barrier in order to cross it, and the probability that a particle visits the barrier is proportional to $e^{-\Delta V}$. However, the results here show that mixing of particles between energy minima can be accelerated by enhancing the probability that if a particle reaches a region with high V, it takes advantage of this excursion in order to cross the barrier. The mechanisms for this

enhanced probability differ between the models considered here—it would be interesting to investigate this effect further, so as to understand how general these mechanisms are and how they can be exploited in practical applications.

4 Conclusion and Outlook

We have considered interacting particle systems described by Markov chains, and their hydrodynamic limits, as described by macroscopic fluctuation theory. We compare reversible and irreversible processes: for an irreversible system with generator $\mathcal{L}$, the corresponding reversible process is the one identified in (7), whose generator is $\mathcal{L}_S$. At the microscopic level, it is known that the irreversible process then converges to its steady state at least as fast as the reversible one—this can be demonstrated by considering either the spectral gap or the (level-2) large deviations of the empirical measure. In the hydrodynamic limit, Eq. (38) shows that this property is preserved, by considering the large deviations of the empirical density. Moreover, Eq. (51) gives a quantitative expression for the acceleration of convergence, which may be seen as a generalisation of previous results for single-particle diffusions [35].

Our numerical results for the ZRP reinforce the observation that for a given reversible system, there is a large family of irreversible systems for which convergence to equilibrium is faster (or, at least, equally fast). We considered two cases: either a drift force in a single direction, which acts to drive a system around a circle (Sect. 3.2.2) or the introduction of a force that drives the system around the level sets of the potential (Sect. 3.2.4). In both cases, we observe acceleration of convergence, as expected.

The results within MFT provide a geometrical interpretation of the acceleration, in terms of forces that act in directions perpendicular to the free energy gradient, as shown by orthogonality relations for currents such as Eq. (46). We have argued that such forces can act to accelerate convergence by driving the system away from regions where the free energy gradient is shallow, in which cases reversible processes exhibit slow convergence.

We offer two perspectives on future application of these ideas. First, we have shown that breaking detailed balance generically accelerates convergence, but of course there are very many ways to write down irreversible models, and it is not clear what choices are most practical in applications, nor which ones lead to the fastest convergence. In particular, the choice considered for ZRP examples shown here are rather specific to systems in one or two dimensions. (We emphasise however that the configuration spaces of the ZRP are very high-dimensional since we consider N interacting particles, so the methods are not restricted to systems with low-dimensional configuration spaces.) Second, we gave a geometrical interpretation in which the symmetric dynamics correspond to the gradient flow (steepest descent) of the free energy and the antisymmetric dynamics are in some sense orthogonal to this gradient flow. This offers a potentially new perspective on hydrodynamic limits in irreversible systems, which it would be interesting to investigate further, for example with a view towards obtaining analytic estimates for the rate of convergence.

Supporting data for this manuscript and the code used for the simulations will be made available short after publication on the University of Bath data archive (DOI:10.15125/BATH-00365).

Acknowledgements MK is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/L015684/1. JZ gratefully acknowledges funding by the EPSRC through project EP/ K027743/1, the Leverhulme Trust (RPG-2013-261) and a Royal Society Wolfson Research Merit Award. This research made use of the Balena High Performance Computing

(HPC) Service at the University of Bath. The authors thank the anonymous referees for very helpful comments and suggestions.

References

1. Adams, S., Dirr, N., Peletier, M., Zimmer, J.: Large deviations and gradient flows. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **371**(2005), 20120341, 17 (2013)
2. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis, Stochastic Modelling and Applied Probability, vol. 57. Springer, New York (2007)
3. Baiesi, M., Maes, C., Wynants, B.: Fluctuations and response of nonequilibrium states. Phys. Rev. Lett. **103**, 010602 (2009)
4. Barato, A.C., Chetrite, R.: A formal view on level 2.5 large deviations and fluctuation relations. J. Stat. Phys. **160**(5), 1154–1172 (2015)
5. Bernard, E.P., Krauth, W., Wilson, D.B.: Event-chain Monte Carlo algorithms for hard-sphere systems. Phys. Rev. E **80**, 056704 (2009)
6. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Fluctuations in stationary nonequilibrium states of irreversible processes. Phys. Rev. Lett. **87**(4), 040601 (2001)
7. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Macroscopic fluctuation theory for stationary non-equilibrium states. J. Stat. Phys. **107**(3–4), 635–675 (2002)
8. Bertini, L., Faggionato, A., Gabrielli, D.: Large deviations of the empirical flow for continuous time Markov chains. Ann. Inst. Henri Poincaré Probab. Stat. **51**(3), 867–900 (2015)
9. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Macroscopic fluctuation theory. Rev. Mod. Phys. **87**(2), 593–636 (2015)
10. Bierkens, J.: Non-reversible Metropolis-Hastings. Stat. Comput. **26**, 1–16 (2015)
11. Bodineau, T., Derrida, B.: Current fluctuations in nonequilibrium diffusive systems: an additivity principle. Phys. Rev. Lett. **92**(18), 180601 (2004)
12. Chen, F., Lovász, L., Pak, I.: Lifting Markov chains to speed up mixing. In: Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999), pp. 275–281 (electronic). ACM, New York (1999)
13. Derrida, B.: Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. J. Stat. Mech. Theory Exp. **2007**(7), P07023 (electronic) (2007)
14. Diaconis, P.: The Markov chain Monte Carlo revolution. Bull. Am. Math. Soc. (N.S.) **46**(2), 179–205 (2009)
15. Diaconis, P., Holmes, S., Neal, R.M.: Analysis of a nonreversible Markov chain sampler. Ann. Appl. Probab. **10**(3), 726–752 (2000)
16. Diaconis, P., Saloff-Coste, L.: Logarithmic Sobolev inequalities for finite Markov chains. Ann. Appl. Probab. **6**(3), 695–750 (1996)
17. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time IV. Commun. Pure Appl. Math. **36**(2), 183–212 (1983)
18. Duncan, A.B., Lelièvre, T., Pavliotis, G.A.: Variance reduction using nonreversible Langevin samplers. J. Stat. Phys. **163**(3), 457–491 (2016)
19. Evans, M.R., Hanney, T.: Nonequilibrium statistical mechanics of the zero-range process and related models. J. Phys. A **38**(19), R195–R240 (2005)
20. Gilbarg, D., Trudinger, N.S.: Elliptic partial differential equations of second order. Class. Math. Springer, Berlin (2001). Reprint of the 1998 edition
21. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**(25), 2340–2361 (1977)
22. Hirschberg, O., Mukamel, D., Schütz, G.M.: Density profiles, dynamics, and condensation in the ZRP conditioned on an atypical current. J. Stat. Mech. Theory Exp. **2015**(11), P11023 (2015)
23. den Hollander, F.: Large Deviations, Fields Institute Monographs, vol. 14. American Mathematical Society, Providence, RI (2000)
24. Hwang, C.R., Hwang-Ma, S.Y., Sheu, S.J.: Accelerating diffusions. Ann. Appl. Probab. **15**(2), 1433–1444 (2005)

25. Hwang, C.R., Normand, R., Wu, S.J.: Variance reduction for diffusions. Stoch. Process. Appl. **125**(9), 3522–3540 (2015)
26. Ichiki, A., Ohzeki, M.: Violation of detailed balance accelerates relaxation. Phys. Rev. E **88**(2), 020101 (2013)
27. Jack, R., Zimmer, J.: Geometrical interpretation of fluctuating hydrodynamics in diffusive systems. J. Phys. A: Math. Theor. **47**, 485001 (2014)
28. Kipnis, C., Landim, C.: Scaling limits of interacting particle systems. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 320. Springer, Berlin (1999)
29. Labbé, C., Lacoin, H.: Cutoff phenomenon for the asymmetric simple exclusion process and the biased card shuffling. arXiv preprint arXiv:1610.07383 (2016)
30. Lelièvre, T., Nier, F., Pavliotis, G.A.: Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. J. Stat. Phys. **152**(2), 237–274 (2013)
31. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov chains and mixing times. American Mathematical Society, Providence, RI (2009). With a chapter by James G. Propp and David B. Wilson
32. Liggett, T.M.: Interacting particle systems. Classics in Mathematics. Springer, Berlin (2005). Reprint of the 1985 original
33. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. **261**(8), 2250–2292 (2011)
34. Newman, M.E.J., Barkema, G.T.: Monte Carlo methods in statistical physics. The Clarendon Press, Oxford University Press, New York (1999)
35. Rey-Bellet, L., Spiliopoulos, K.: Irreversible Langevin samplers and variance reduction: a large deviations approach. Nonlinearity **28**(7), 2081–2103 (2015)
36. Rey-Bellet, L., Spiliopoulos, K.: Improving the convergence of reversible samplers. J. Stat. Phys. **164**(3), 472–494 (2016)
37. Sakai, Y., Hukushima, K.: Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions. Phys. Rev. E **93**(4), 043318 (2016)
38. Spitzer, F.: Interaction of Markov processes. Adv. Math. **5**, 246–290 (1970)
39. Sun, Y., Schmidhuber, J., Gomez, F.J.: Improving the asymptotic performance of Markov Chain Monte-Carlo by inserting vortices. In: Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R.S., Culotta, A. (eds.) Advances in Neural Information Processing Systems 23, pp. 2235–2243. Neural Information Processing Systems Foundation (2010)
40. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. **478**(1–3), 1–69 (2009)

2.2. Conclusions

We presented several arguments why irreversible processes generically converge faster than reversible processes to their steady state, both on the microscopic (Markov chain) scale and the macroscopic scale. We put a special emphasis on the hydrodynamic scale, where we obtained two results of interest.

The first result is Equation (51) in Proposition 5, which is based on a splitting of the hydrodynamic current in the sum of three separate parts that are orthogonal in the sense of Equation (30) (cf. Proposition 4). There we obtained a level-2 like rate functional, which characterises the exponential decay rate for the probability to observe a time averaged empirical density $\frac{1}{T}\int_0^T \rho_s^L \mathrm{d}s$, which differs from the steady state $\bar{\rho}$, as first $L \to \infty$ and then $T \to \infty$. Our finding is that the reversible rate functional is dominated by the irreversible one

$$I^{\mathrm{rev}}(\rho) \leq I^{\mathrm{irr}}(\rho).$$

Similar to [46], where this criterion was used to argue that irreversible diffusions converge faster than reversible ones, we can interpret this result as an acceleration of convergence.

Our second result is based on the fact that J_S, the symmetric part of the dynamics J, can be related to a gradient flow for the quasipotential $\mathcal{V}$ (cf. Equation (29)) in a generalised Wassertein metric [24]. Combined with the splitting (30), this yields a geometric interpretation of the dynamics on the space of measures. More precisely, (30) allows to show that the quasipotential is invariant under the anti-symmetric part of the dynamics J_A, inducing an evolution on the level sets of $\mathcal{V}$, see Equation (31). We argued that the relaxation for the reversible (steepest descent) dynamics can be slow when the gradient of $\mathcal{V}$ is shallow, in which case the anti-symmetric part of the dynamics can drive the system away from these shallow areas.

For Markov chains, the quantity corresponding to the quasipotential $\mathcal{V}$ is the free energy $\mathcal{F}$, which coincides with the relative entropy (or Kullback-Leibler divergence). As opposed to the case above, one can show that the anti-symmetric part of the dynamics does not leave the free energy invariant. The latter can be seen to be related to the question if and how one can define an appropriate structure for Markov chains that allows for a generalisation of Equation (30). One would like to define a 'generalised geometry', which includes both, Markov chains and MFT, that (e.g.) explains why the quasipotential is invariant under the irreversible dynamics, whereas the free energy for Markov chains is *not*. Finally, one should be able to relate this structure to Large Deviation Principles (LDPs), as it is the case in MFT (see Equation (25)). We will answer this question in the following chapter, where we introduce a canonical structure of Markov chains.

Chapter 3

Canonical Structure and Orthogonality of Forces and Currents in Irreversible Markov Chains

In this chapter we discuss a variational structure for fluctuations in Markov chains, which was originally derived by Maes and co-workers (who first referred to this structure as 'canonical structure'), see [39, 40], which can be related to a Ψ-$\Psi^\star$ structure (see e.g. the work by Mielke et al. [41, 42]). We present a novel splitting, which generalises (30) from Chapter 2. The findings presented in this chapter have been published by the author of this thesis, Robert L. Jack and Johannes Zimmer in the *Journal of Statistical Physics* as open access publication [29].

3.1. Outline of the Article

Within Macroscopic Fluctuation Theory, the *local Einstein relation* (see Equation (34) in Chapter 2) allows to relate the current $J(\rho)$ to a force $F(\rho)$ acting on the system. They satisfy a linear flux-force relation, in the sense that $J(\rho) = \chi(\rho)F(\rho)$ for the mobility matrix $\chi(\rho)$. This linear flux-force relation can be shown to be equivalent to a quadratic structure of the *Onsager-Machlup functional*, which is in this case (for a current j) given by

$$\Phi_{\mathrm{MFT}}(\rho, j, F(\rho)) = \frac{1}{2}\int_\Lambda (j - \chi(\rho)F(\rho)) \cdot \chi(\rho)^{-1}(j - \chi(\rho)F(\rho))\,\mathrm{d}u.$$

Note that Φ_{MFT} appeared in the large deviation rate function for dynamical large deviations (25) in Chapter 2 and also in the rate functions (37) and (44) in Chapter 2.

For Markov chains, we can consider a probability current $J(\rho)$, where ρ is the probability distribution of the Markov chain. We remark that the current $J(\rho)$ replaces the generator $\mathcal{L}$ considered in Chapter 2. In this chapter we show that Markov chains satisfy a non-linear flux-force relation of the form $J(\rho) = a(\rho)\sinh(F(\rho)/2)$, where $a(\rho)$ corresponds to the mobility and $F(\rho)$ is the force acting on the system. Consequently, the associated Onsager-Machlup functional is a non-quadratic functional, which can be shown to be of the form

$$\Phi(\rho, j, f) = \Psi(\rho, j) - j \cdot f + \Psi^\star(\rho, f),$$

where Ψ and $\Psi^\star$ are Legendre dual w.r.t. a dual pairing $j \cdot f$ (see Section 2.2). One then can show that the rate function for dynamical fluctuations (24), as well as the level-2.5 and level-2 rate functions can be stated in terms of Φ (see e.g. Equations (64) and (81) below). We further explain how the force $F(\rho)$ can be split in a time-reversal symmetric part $F^S(\rho)$ and an anti-symmetric part F^A (which does not depend on ρ). In physics, these two forces have interpretations as *entropy production* and *housekeeping heat*, respectively (see Section 3.1 and Section 3.2). In Proposition 3, we then prove a 'generalised orthogonality' for $F^S(\rho)$ and F^A, which can be interpreted as a generalisation of (30) in Chapter 2. Finally, we show in Section 6 how many results obtained in MFT can be obtained for general processes with a Ψ-$\Psi^\star$ structure (where MFT and Markov chains are two examples).

Appendix B: Statement of Authorship

This declaration concerns the article entitled:									
Canonical Structure and Orthogonality of Forces and Currents in Irreversible Markov Chains									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**		**Accepted**		**Published**	X
Publication details (reference)	Journal: Journal of Statistical Physics, DOI: 10.1007/s10955-018-1986-0 Authors: Marcus Kaiser, Robert L. Jack and Johannes Zimmer								
Candidate's contribution to the paper (detailed, and also given as a percentage).	The bulk of the calculations have been performed by the author of the thesis (70%). All authors contributed equally to the presentation of the content (33%).								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed					**Date**	10.7.2018			

J Stat Phys (2018) 170:1019–1050
https://doi.org/10.1007/s10955-018-1986-0

Canonical Structure and Orthogonality of Forces and Currents in Irreversible Markov Chains

Marcus Kaiser[1] · **Robert L. Jack**[2,3,4] · **Johannes Zimmer**[1]

Received: 4 August 2017 / Accepted: 6 February 2018 / Published online: 15 February 2018

Abstract We discuss a canonical structure that provides a unifying description of dynamical large deviations for irreversible finite state Markov chains (continuous time), Onsager theory, and Macroscopic Fluctuation Theory (MFT). For Markov chains, this theory involves a nonlinear relation between probability currents and their conjugate forces. Within this framework, we show how the forces can be split into two components, which are orthogonal to each other, in a generalised sense. This splitting allows a decomposition of the pathwise rate function into three terms, which have physical interpretations in terms of dissipation and convergence to equilibrium. Similar decompositions hold for rate functions at level 2 and level 2.5. These results clarify how bounds on entropy production and fluctuation theorems emerge from the underlying dynamical rules. We discuss how these results for Markov chains are related to similar structures within MFT, which describes hydrodynamic limits of such microscopic models.

Keywords Nonequilibrium dynamical fluctuations · Large deviations · Microscopic fluctuation theory · Irreversible Markov chains

Mathematics Subject Classification 82C22 · 82C35 · 60J27 · 60F10

Marcus Kaiser
m.kaiser@bath.ac.uk

Robert L. Jack
rlj22@cam.ac.uk

Johannes Zimmer
j.zimmer@bath.ac.uk

1 Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

2 Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

3 Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

4 Department of Physics, University of Bath, Bath BA2 7AY, UK

1 Introduction

We consider dynamical fluctuations in systems described by Markov chains. The nature of such fluctuations in physical systems constrains the mathematical models that can be used to describe them. For example, there are well-known relationships between equilibrium physical systems and detailed balance in Markov models [20, Sect. 5.3.4]. Away from equilibrium, fluctuation theorems [12,19,25,32,37] and associated ideas of local detailed balance [32, 39] have shown how the entropy production of a system must be accounted for correctly when modelling physical systems. However, the mathematical structures that determine the probabilities of non-equilibrium fluctuations are still only partially understood.

We characterise dynamical fluctuations using an approach based on the *Onsager–Machlup (OM) theory* [36], which is concerned with fluctuations of macroscopic properties of physical systems (for example, density or energy). Associated to these fluctuations is a *large-deviation principle* (LDP), which encodes the probability of rare dynamical trajectories. The classical ideas of OM theory have been extended in recent years, through the *Macroscopic Fluctuation Theory* (MFT) of Bertini et al. [7]. This theory uses an LDP to describe path probabilities for the density and current in diffusive systems, on the hydrodynamic scale. At the centre of MFT is a decomposition of the current into two orthogonal terms, one of which is symmetric under time-reversal, and another which is anti-symmetric. The resulting theory is a general framework for the analysis of dynamical fluctuations in a large class of non-equilibrium systems. It also connects dynamical fluctuations with thermodynamic quantities like free energy and entropy production, and with associated non-equilibrium objects like the quasi-potential (which extends the thermodynamic free energy to non-equilibrium settings).

Here, we show how several features that appear in MFT can be attributed to a general structure that characterises dynamical fluctuations in microscopic Markov models. That is, the properties of the hydrodynamic (MFT) theory can be traced back to the properties of the underlying stochastic processes. Our approach builds on recent work by Mielke, Renger and M. A. Peletier, in which the analogue of the OM theory for reversible Markov chains has been described in terms of a *generalised gradient-flow structure* [43]. To describe non-equilibrium processes, that theory must be generalised to include irreversible Markov chains. This can be achieved using the canonical structure of fluctuations discovered by Maes and Netočný [38]. Extending their approach, we decompose currents in the system into two parts, and we identify a kind of orthogonality relationship associated with this decomposition. However, in contrast to the classical OM theory and to MFT, the large deviation principles that appear in our approach have non-quadratic rate functions, which means that fluxes have non-linear dependence on their conjugate forces. Thus, the idea of orthogonality between currents needs to be generalised, just as the notion of gradient flows in macroscopic equilibrium systems can be extended to generalised gradient flows.

The central players in our analysis are the probability density ρ and the probability current j. For a given Markov chain, the relation between these quantities is fully encoded in the master equation, which also fully specifies the dynamical fluctuations in that model. However, thermodynamic aspects of the system—the roles of heat, free energy, and entropy production—are not apparent in the master equation. Within the Onsager–Machlup theory, these thermodynamic quantities appear in the action functional for paths, and solutions of the master equation appear as paths of minimal action. Hence, the structure that we discuss here, and particularly the decomposition of the current into two components, links the dynamical properties of the system to thermodynamic concepts, both for equilibrium and non-equilibrium systems.

1.1 Summary

We now sketch the setting considered in this article (precise definitions of the systems of interest and the relevant currents, densities and forces will be given in Sect. 2).

We introduce a large parameter $\mathcal{N}$, which might be the size of the system (as in MFT) or a large number of copies of the system (an ensemble), as considered for Markov chains in [39]. Then let $(\hat{\rho}_t^{\mathcal{N}}, \hat{j}_t^{\mathcal{N}})_{t\in[0,T]}$ be the (random) path followed by the system's density and current, in the time interval $[0, T]$. Consider a random initial condition such that $\text{Prob}(\hat{\rho}_0^{\mathcal{N}} \approx \rho) \asymp \exp[-\mathcal{N} I_0(\rho)]$, asymptotically as $\mathcal{N} \to \infty$, for some rate functional I_0. Paths that in addition satisfy a continuity equation $\dot{\rho} + \text{div}\, j = 0$ have the asymptotic probability

$$\text{Prob}\left(\left(\hat{\rho}_t^{\mathcal{N}}, \hat{j}_t^{\mathcal{N}}\right)_{t\in[0,T]} \approx (\rho_t, j_t)_{t\in[0,T]}\right) \asymp \exp\left\{-\mathcal{N} I_{[0,T]}\left((\rho_t, j_t)_{t\in[0,T]}\right)\right\} \tag{1}$$

with the *rate functional*

$$I_{[0,T]}\big((\rho_t, j_t)_{t\in[0,T]}\big) = I_0(\rho_0) + \frac{1}{2}\int_0^T \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t; \tag{2}$$

here $F(\rho_t)$ is a force (see (12) below for the precise definition) and Φ is what we call the *generalised OM functional*, which has the general form

$$\Phi(\rho, j, f) := \Psi(\rho, j) - j \cdot f + \Psi^\star(\rho, f), \tag{3}$$

where $j \cdot f$ is a dual pairing between a current j and a force f, while Ψ and $\Psi^\star$ are a pair of functions which satisfy

$$\Psi^\star(\rho, f) = \sup_j\big[j \cdot f - \Psi(\rho, j)\big], \quad \text{and} \quad \Psi(\rho, j) = \sup_f\big[j \cdot f - \Psi^\star(\rho, f)\big], \tag{4}$$

as well as $\Psi^\star(\rho, f) = \Psi^\star(\rho, -f)$ and $\Psi(\rho, j) = \Psi(\rho, -j)$. Note that (4) means that the two functions satisfy a Legendre duality. Moreover, these two functions Ψ and $\Psi^\star$ are strictly convex in their second arguments. Here and throughout, f indicates a force, while F is a function whose (density-dependent) value is a force.

The large deviation principle stated in (1) is somewhat abstract: for example, $\hat{\rho}_t^{\mathcal{N}}$ might be defined as a density on a discrete space or on $\mathbb{R}^d$, depending on the system of interest. Specific examples will be given below. In addition, all microscopic parameters of the system (particle hopping rates, diffusion constants, etc.) will enter the (system-dependent) functions Ψ, $\Psi^\star$ and F.

As a preliminary example, we recall the classical Onsager theory [36], in which one considers n currents $j = (j^\alpha)_{\alpha=1}^n$ and a set of conjugate applied forces $F = (F^\alpha)_{\alpha=1}^n$. Examples of currents might be particle flow or heat flow, and the relevant forces might be pressure or temperature gradients. The large parameter $\mathcal{N}$ corresponds to the size of a macroscopic system. The theory aims to to describe the typical (average) response of the current j to the force F, and also the fluctuations of j. In this (simplest) case, the density ρ plays no role, so the force F has a fixed value in $\mathbb{R}^n$. The dual pairing is simply $j \cdot f = \sum_\alpha j^\alpha f^\alpha$ and Ψ is given by $\Psi(\rho, j) = \frac{1}{2}\sum_{\alpha,\beta} j^\alpha R^{\alpha\beta} j^\beta$, where R is a symmetric $n \times n$ matrix with elements $R^{\alpha\beta}$. The Legendre dual of Ψ is $\Psi^\star(\rho, f) = \frac{1}{2}\sum_{\alpha,\beta} f^\alpha L^{\alpha\beta} f^\beta$, where $L = R^{-1}$ is the *Onsager matrix*, whose elements are the linear response coefficients of the system. One sees that Ψ and $\Psi^\star$ can be interpreted as squared norms for currents and forces respectively. Denoting this norm by $\|j\|_{L^{-1}}^2 := \Psi(\rho, j)$, one has

$$\Phi(\rho, j, f) = \|j - Lf\|_{L^{-1}}^2. \tag{5}$$

On applying an external force F, the response of the current j is obtained as the minimum of Φ, so $j = LF$ (that is, $j^\alpha = \sum_\beta L^{\alpha\beta} F^\beta$). One sees that Φ measures the deviation of the current j from its expected value LF, within an appropriate norm. From the LDP (1), one sees that the size of this deviation determines the probability of observing a current fluctuation of this size.

In this article, we show in Sect. 2 that finite Markov chains have an LDP rate functional of the form (3), where Φ (and thus $\Psi^\star$) are *not* quadratic. In that case, ρ and j correspond to probability densities and probability currents, while the transition rates of the Markov chain determine the functions F, Ψ and $\Psi^\star$. Since Ψ and $\Psi^\star$ measure respectively the sizes of the currents and forces, we interpret them as generalisations of the squared norms that appear in the classical case. The resulting Φ is not a squared norm, but it is still a non-negative function that measures the deviation of j from its most likely value. This leads to nonlinear relations between forces and currents. The MFT theory [7] also fits in this framework, as we show in Sect. 4: in that case ρ, j are a particle density and a particle current. However, there are relationships between the functions Φ for MFT and for general Markov chains, as we discuss in Sect. 4.5.

Hence, the general structure of Eqs. (1)–(4) describes classical OM theory [36], MFT, and finite Markov chains. A benefit is that the terms have a physical interpretation. For a path (ρ, j), the time-reversed path is $(\rho_t^*, j_t^*) := (\rho_{T-t}, -j_{T-t})$. Since both Ψ and $\Psi^\star$ are symmetric in their second argument and thus invariant under time reversal, it holds that $\Phi(\rho, j, f) - \Phi(\rho^*, j^*, f) = -2j \cdot f$. This allows us to identify $j \cdot F(\rho)$ as a rate of entropy production. In contrast, the term $\Psi(\rho, j) + \Psi^\star(\rho, F(\rho))$ is symmetric under time reversal and encodes the frenesy (see [3]). Thus, within this general structure, the physical significance of Eqs. (1)–(4) is that they connect path probabilities to physical notions such as force, current, entropy production and breaking of time-reversal symmetry. Furthermore, we introduce in Sect. 3 decompositions of forces and the (path-wise) rate functional. Sect. 4 shows that some results of MFT originate from generalised orthogonalities of the underlying Markov chains derived in Sect. 3. Similar results hold for time-average large deviation principles, as shown in Sect. 5. In Sect. 6, we show how some properties of MFT can be derived directly from the canonical structure (1)–(4), independent of the specific models of interest. Hence these results of MFT have analogues in Markov chains. Finally we briefly summarise our conclusions in Sect. 7.

2 Onsager–Machlup Theory for Markov Chains

In this section, we collect results on forces and currents in Markov chains and on associated LDPs. In particular, we recall the setting of [38,39]; other references for this section are for example [49] (for the definition of forces and currents in Markov chains) and [43] for LDPs.

2.1 Setting

We consider an irreducible continuous time Markov chain X_t on a finite state space V with a unique stationary distribution π that satisfies $\pi(x) > 0$ for all $x \in V$. The transition rate from state x to state y is denoted with r_{xy}. We assume that $r_{xy} > 0$ if and only if $r_{yx} > 0$.

We restrict to finite Markov chains for simplicity: the theory can be extended to countable state Markov chains, but this requires some additional assumptions. Briefly, one requires that the Markov chain should be positively recurrent and ergodic (see for instance [9]), for which it is sufficient that (i) the transition rates are not degenerate: $\sum_{y \in V} r_{xy} < \infty$ for all $x \in V$,

and (ii) for each $x \in V$, the Markov chain started in x almost all trajectories of the Markov chain do not exhibit infinitely many jumps in finite time ("no explosion"). Second, one has to invoke a summability condition for the currents considered below (see, e.g., Eqs. 9 and 10), such that in particular the discrete integration by parts (or summation by parts) formula (15) holds. Finally, note that the cited result for existence and uniqueness of the optimal control potential (the solution to (70)) is only valid for finite state Markov chains.

As usual, we can interpret the state space of the Markov chain as a directed graph with vertices V and edges $E = \{xy \mid x, y \in V, r_{xy} > 0\}$, such that $xy \in E$ if and only if $yx \in E$. Let ρ be a probability measure on V. We define rescaled transition rates with respect to π as

$$q_{xy} := \pi(x) r_{xy}, \tag{6}$$

so that $\rho(x) r_{xy} = \frac{\rho(x)}{\pi(x)} q_{xy}$. With this notation, the *detailed balance* condition $\pi(x) r_{xy} = \pi(y) r_{yx}$ reads $q_{xy} = q_{yx}$, so this equality holds precisely if the Markov chain is reversible (i.e. satisfies detailed balance). In general (not assuming reversibility), since π is the invariant measure for the Markov chain, one has (for all x) that

$$\sum_y (q_{xy} - q_{yx}) = 0. \tag{7}$$

We further define the *free energy* $\mathcal{F}$ on V to be the *relative entropy* (or *Kullback–Leibler divergence*) with respect to π,

$$\mathcal{F}(\rho) := \sum_x \rho(x) \log\Big(\frac{\rho(x)}{\pi(x)}\Big). \tag{8}$$

The *probability current* $J(\rho)$ is defined as [49, Eq. (7.4)]

$$J_{xy}(\rho) := \rho(x) r_{xy} - \rho(y) r_{yx}. \tag{9}$$

Moreover, for a general current j such that $j_{xy} = -j_{yx}$, we define the *divergence* as

$$\operatorname{div} j(x) := \sum_{y \in V} j_{xy}. \tag{10}$$

We say that j is *divergence free* if $\operatorname{div} j(x) = 0$ for every $x \in V$. The time evolution of the probability density ρ is then given by the master equation

$$\dot{\rho}_t = -\operatorname{div} J(\rho_t) \tag{11}$$

(which is often stated as $\dot{\rho}_t = \mathcal{L}^\dagger \rho_t$, with the (forward) generator $\mathcal{L}^\dagger$).

2.2 Non-linear Flux–Force Relation and the Associated Functionals Ψ and $\Psi^\star$

To apply the theory outlined in Sect. 1.1, the next step is to identify the appropriate forces $F(\rho)$ and also a set of mobilities $a(\rho)$. In this section we define these forces, following [38,39,49]. This amounts to a reparameterisation of the rates of the Markov process in terms of physically-relevant variables: an example is given in Sect. 3.5.

To each edge in E we assign a *force* F and a *mobility* a, as

$$F_{xy}(\rho) := \log \frac{\rho(x) r_{xy}}{\rho(y) r_{yx}} \quad \text{and} \quad a_{xy}(\rho) := 2\sqrt{\rho(x) r_{xy} \rho(y) r_{yx}}. \tag{12}$$

Note that $F_{xy} = -F_{yx}$, while $a_{xy} = a_{yx}$: forces have a direction but the mobility is a symmetric property of each edge. The fact that F_{xy} depends on the density ρ means that these

forces act in the space of probability distributions. This definition of the force is sometimes also called *affinity* [49, Eq. (7.5)]; see also [1]. With this definition, the probability current (9) is

$$J_{xy}(\rho) = a_{xy}(\rho)\sinh\big(\tfrac{1}{2}F_{xy}(\rho)\big), \tag{13}$$

which may be verified directly from the definition $\sinh(x) = (\mathrm{e}^{x} - \mathrm{e}^{-x})/2$. In contrast to the classical OM theory, this is a *non-linear* relation between forces and fluxes, although one recovers a linear structure for small forces (recall the classical theory in Sect. 1.1, for which $j = Lf$).

Now consider a current j defined on E, with $j_{xy} = -j_{yx}$, and a general force f that satisfies $f_{xy} = -f_{yx}$ (which is not in general given by (12)). Define a dual pair on E as

$$j \cdot f := \frac{1}{2}\sum_{xy} j_{xy} f_{xy}, \tag{14}$$

where the summation is over all $xy \in E$ (the normalisation $1/2$ appears because each connected pair of states should be counted only once, but E is a set of directed edges, so it contains both xy and yx, which have the same contribution to $j \cdot f$).

We define the discrete gradient ∇g by $\nabla^{x,y} g := g(y) - g(x)$. The discrete gradient and the divergence defined in (10) satisfy a discrete integration by parts formula: for any function $g\colon V \to \mathbb{R}$, since $j_{xy} = -j_{yx}$, we have

$$-\sum_{x\in V} g(x)\,\mathrm{div}\, j(x) = \frac{1}{2}\sum_{xy} j_{xy}\nabla^{x,y} g = j \cdot \nabla g. \tag{15}$$

We will show in Sect. 2.3 that there is an OM functional associated with these forces and currents, which is of the form (3). Since Ψ and $\Psi^\star$ are convex and related by a Legendre transformation, it is sufficient to specify only one of them. The appropriate choice turns out to be

$$\Psi^\star(\rho, f) := \sum_{xy} a_{xy}(\rho)\big(\cosh\big(\tfrac{1}{2}f_{xy}\big) - 1\big). \tag{16}$$

This means that $\Phi(\rho, j, f)$ defined in (3) is uniquely minimised for the current $j_{xy} = j^f_{xy}(\rho)$ with

$$j^f_{xy}(\rho) = 2(\delta\Psi^\star/\delta f)_{xy} = a_{xy}(\rho)\sinh(f_{xy}/2), \tag{17}$$

as required for consistency with (13). From (4) and (14), one has also

$$\Psi(\rho, j) = \frac{1}{2}\sum_{xy} j_{xy} f^j_{xy}(\rho) - \sum_{xy} a_{xy}(\rho)\big(\cosh\big(\tfrac{1}{2}f^j_{xy}(\rho)\big) - 1\big), \tag{18}$$

where

$$f^j_{xy}(\rho) := 2\,\mathrm{arcsinh}\,\big(j_{xy}/a_{xy}(\rho)\big) \tag{19}$$

is the force required to induce the current j.

Physically, $\Psi^\star(\rho, f)$ is a measure of the strength of the force f and $\Psi(\rho, j)$ is a measure of the magnitude of the current j. Consistent with this interpretation, note that Ψ and $\Psi^\star$ are symmetric in their second arguments. Moreover, for small forces and currents, $\Psi^\star$ and Ψ are quadratic in their second arguments, and can be interpreted as generalisations of squared norms of the force and current respectively. Note that Eqs. (16) and (18) can alternatively be represented as

$$\Psi(\rho, j) = \sum_{xy}\left[\frac{1}{2} j_{xy} f^j_{xy}(\rho) - \sqrt{j_{xy}^2 + a_{xy}(\rho)^2} + a_{xy}(\rho)\right] \tag{20}$$

and

$$\Psi^\star(\rho, f) := \sum_{xy}\left[\sqrt{j^f_{xy}(\rho)^2 + a_{xy}(\rho)^2} - a_{xy}(\rho)\right]. \tag{21}$$

2.3 Large Deviations and the Onsager–Machlup Functional

As anticipated in Sect. 1.1, the motivation for the definitions of Ψ, $\Psi^\star$, and F is that there is a large deviation principle for these Markov chains, whose rate function is of the form given in (2). This large deviation principle appears when one considers $\mathcal{N}$ independent copies of the Markov chain.

We denote the ith copy of the Markov chain by X^i_t and define the empirical density for this copy as $\hat\rho^i_t(x) = \delta_{X^i_t, x}$, where δ is a Kronecker delta function. Let the times at which the Markov chain X^i_t has jumps in $[0, T]$ be $t^i_1, t^i_2, \ldots, t^i_{K_i}$. Further denote the state just before the kth jump with x^i_{k-1} (such that the state after the kth jump is x^i_k). With this, the empirical current is given by

$$\left(\hat\jmath^i_t\right)_{xy} = \sum_{k=1}^{K_i}\left(\delta_{x,x^i_{k-1}}\delta_{y,x^i_k} - \delta_{y,x^i_{k-1}}\delta_{x,x^i_k}\right)\delta\left(t - t^i_k\right),$$

where $\delta(t - t_k)$ denotes a Dirac delta. Note that $(\hat\jmath^i_t)_{xy} = -(\hat\jmath^i_t)_{yx}$ and the total probability is conserved, as $\sum_x \operatorname{div} \hat\jmath^i_t(x) = 0$ (which holds for any discrete vector field with $(\hat\jmath^i_t)_{xy} = -(\hat\jmath^i_t)_{yx}$). With a slight abuse of notation we define a similar empirical density and current for the full set of copies as

$$\hat\rho^{\mathcal{N}}_t := \frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}} \hat\rho^i_t, \quad \text{and} \quad \hat\jmath^{\mathcal{N}}_t := \frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}} \hat\jmath^i_t. \tag{22}$$

Next, we state the large deviation principle where the OM functional appears. For this, we fix a time interval $[0, T]$ and consider the large $\mathcal{N}$ limit. We assume that the $\mathcal{N}$ copies at time $t = 0$ have initial conditions drawn from the invariant measure of the process (the generalisation to other initial conditions is straightforward). Then, the probability to observe a joint density and current $(\rho_t, j_t)_{t\in[0,T]}$ over the time interval $[0, T]$ is in the limit as $\mathcal{N} \to \infty$ given by (1). That is,

$$\operatorname{Prob}\left(\left(\hat\rho^{\mathcal{N}}_t, \hat\jmath^{\mathcal{N}}_t\right)_{t\in[0,T]} \approx (\rho_t, j_t)_{t\in[0,T]}\right) \asymp \exp\left\{-\mathcal{N} I_{[0,T]}\left((\rho_t, j_t)_{t\in[0,T]}\right)\right\} \tag{23}$$

with

$$I_{[0,T]}\left((\rho_t, j_t)_{t\in[0,T]}\right) = \begin{cases} \mathcal{F}(\rho_0) + \frac{1}{2}\int_0^T \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t & \text{if } \dot\rho_t + \operatorname{div} j_t = 0 \\ +\infty & \text{otherwise} \end{cases} \tag{24}$$

Here, $F(\rho)$ is the force defined in (12) and the condition $\dot\rho_t + \operatorname{div} j_t = 0$ has to hold for almost all $t \in [0, T]$. Moreover, Φ is of the form $\Phi(\rho, j, f) = \Psi(\rho, j) - j \cdot f + \Psi^\star(\rho, f)$ stated in (3), and the relevant functions Ψ, $\Psi^\star$ and $\mathcal{F}$ are those of (16), (18) and (8). This LDP was formally derived in [38,39]. Since the quantities defined in (22) are simple averages over independent copies of the same Markov chain, this LDP may also be proven by direct

application of Sanov's theorem, which provides an interpretation of $I_{[0,T]}$ as a relative entropy between path measures; we sketch the derivation in Appendix A. For finite-state Markov chains, (23) and (24) also follow (by contraction) from [48, Theorem 4.2], which provides a rigorous proof.

We emphasise that the arguments ρ and j of the function Φ correspond to the random variables that appear in the LDP, while the functions F, Ψ and $\Psi^\star$ that appear in Φ encapsulate the transition rates of the Markov chain. Thus, by reparameterising the rates r_{xy} in terms of forces F and mobilities a, we arrive at a representation of the rate function which helps to make its properties transparent (convexity, positivity, symmetries such as (25)).

We note that for reversible Markov chains, the force $F(\rho)$ is a pure gradient $F = \nabla G$ for some potential G (see Sect. 3), in which case one may write $j \cdot F = \sum_x \dot\rho(x) G(x)$, which follows from an integration by parts and application of the continuity equation. In this case, Mielke, M. A. Peletier, and Renger [43] also identified a slightly different canonical structure to the one presented here, in which the dual pairing is $\sum_x v(x)G(x)$, for a velocity $v(x) = \dot\rho(x)$ and a potential G. The analogues of Ψ and $\Psi^\star$ in that setting depend on v and G respectively, instead of j and F. The setting of (3) and (4) is more general, in that the functions $\Psi, \Psi^\star$ for the velocity/potential setting are fully determined by those for the current/force setting. Also, focusing on the velocity v prevents any analysis of the divergence-free part of the current, and restricting to potential forces does not generalise in a simple way to irreversible Markov chains. For this reason, we use the current/force setting in this work.

In a separate development, Maas [35] identified a quadratic cost function for paths (in fact a metric structure) for which the master equation (11) is the minimiser in the case of reversible dynamics. This metric corresponds to the solution of an optimal mass transfer problem which seems to have no straightforward extension to irreversible systems. Of course, in the reversible case, the pathwise rate function (24) has the same minimiser, but is non-quadratic and therefore does not correspond to a metric structure, so there is no simple geometrical interpretation of (24). It seems that the non-quadratic structure in the rate function is essential in order capture the large deviations encoded by (23).

2.4 Time-Reversal Symmetry, Entropy Production, and the Gallavotti–Cohen Theorem

The rate function for the large-deviation principle (23) is given by (24), which has been written in terms of forces F, currents j, and densities ρ. To explain why it is useful to write the rate function in this way, we compare the probability of a path $(\rho_t, j_t)_{t\in[0,T]}$ with that of its time-reversed counterpart $(\rho_t^*, j_t^*)_{t\in[0,T]}$, where $(\rho_t^*, j_t^*) = (\rho_{T-t}, -j_{T-t})$ as before.

In this case, the fact that Ψ and $\Psi^\star$ are both even in their second argument means that

$$\begin{aligned} -\frac{1}{\mathcal{N}} \log \frac{\mathrm{Prob}\Big(\big(\hat\rho_t^{\mathcal{N}}, \hat j_t^{\mathcal{N}}\big)_{t\in[0,T]} \approx (\rho_t, j_t)_{t\in[0,T]}\Big)}{\mathrm{Prob}\Big(\big(\hat\rho_t^{\mathcal{N}}, \hat j_t^{\mathcal{N}}\big)_{t\in[0,T]} \approx \big(\rho_t^*, j_t^*\big)_{t\in[0,T]}\Big)} \\ \asymp I_{[0,T]}\Big((\rho_t, j_t)_{t\in[0,T]}\Big) - I_{[0,T]}\Big(\big(\rho_t^*, j_t^*\big)_{t\in[0,T]}\Big) \\ = \mathcal{F}(\rho_0) - \mathcal{F}(\rho_T) - \int_0^T j_t \cdot F(\rho_t)\,\mathrm{d}t. \end{aligned} \tag{25}$$

This formula is a (finite-time) statement of the Gallavotti–Cohen fluctuation theorem [19,32]: see also [12,37]. It also provides a connection to physical properties of the system being modelled, via the theory of stochastic thermodynamics [50]. The terms involving the free

energy $\mathcal{F}$ come from the initial conditions of the forward and reverse paths, while the integral of $j \cdot F$ corresponds to the heat transferred from the system to its environment during the trajectory [50, Eqs. (18), (20)]. This latter quantity—which is the time-reversal antisymmetric part of the pathwise rate function—is related (by a factor of the environmental temperature) to the entropy production in the environment [37]. The definition of the force F in (12) has been chosen so that the dual pairing $j \cdot F$ is equal to this rate of heat flow: this means that the forces and currents are conjugate variables, just as (for example) pressure and volume are conjugate in equilibrium thermodynamics. See also the example in Sect. 3.5.

3 Decomposition of Forces and Rate Functional

We now introduce a splitting of the force $F(\rho)$ into two parts $F^S(\rho)$ and F^A, which are related to the behaviour of the system under time-reversal, as well as to the splitting of the heat current into "excess" and "housekeeping" contributions [50]. We use this splitting to decompose the function Φ into three pieces, which allows us to compare (for example) the behaviour of reversible and irreversible Markov chains. This splitting also mirrors a similar construction within MFT [7], and this link will be discussed in Sect. 4. Related splittings have been introduced elsewhere; see [30] and [47] for decompositions of forces in stochastic differential equations, and [13] for decompositions of the instantaneous current in interacting particle systems.

3.1 Splitting of the Force According to Time-Reversal Symmetry

We define the *adjoint process* associated with the original Markov chain of interest. The transition rates of the adjoint process are $r^*_{xy} := \pi(y) r_{yx} \pi(x)^{-1}$. It is easily verified that the adjoint process has invariant measure π, so $q^*_{xy} := \pi(x) r^*_{xy} = q_{yx}$. Under the assumption that the initial distribution is sampled from the steady state, the probability to observe a trajectory for the adjoint process coincides with the probability to observe the time-reversed trajectory for the original process.

From the definition of $F(\rho)$ in (12), we can decompose this force as

$$F_{xy}(\rho) = F^S_{xy}(\rho) + F^A_{xy} \tag{26}$$

with

$$F^S_{xy}(\rho) := -\nabla^{x,y} \log \frac{\rho}{\pi}, \qquad F^A_{xy} := \log \frac{q_{xy}}{q_{yx}}. \tag{27}$$

With this choice, we note that the equivalent force for the adjoint process

$$F^*(\rho) = \log \frac{\rho(x) r^*_{xy}}{\rho(y) r^*_{yx}},$$

satisfies $F^*(\rho) = F^S(\rho) - F^A$. So taking the adjoint inverts the sign of F^A (the "antisymmetric" force) but leaves $F^S(\rho)$ unchanged (the "symmetric" force). For a reversible Markov chain, the adjoint process coincides with the original one, and $F^A = 0$.

Lemma 1 *Given ρ, with the mobility $a(\rho)$ of (12), the forces $F^S(\rho)$ and F^A satisfy*

$$\sum_{xy} \sinh\Big(F^S_{xy}(\rho)/2\Big)\, a_{xy}(\rho) \sinh\Big(F^A_{xy}/2\Big) = 0. \tag{28}$$

Proof From the definitions of $F^S(\rho)$, F^A, a_{xy} and sinh, one has

$$a_{xy}(\rho)\sinh\left(F^S_{xy}(\rho)/2\right) = \left(\frac{\rho(x)}{\pi(x)} - \frac{\rho(y)}{\pi(y)}\right)\sqrt{q_{xy}q_{yx}}$$

and $\sinh(F^A_{xy}/2) = (q_{xy}q_{yx})^{-1/2}(q_{xy} - q_{yx})/2$. Hence

$$\begin{aligned}\sum_{xy}\sinh\Big(F^S_{xy}(\rho)/2\Big)a_{x,y}(\rho)\sinh\Big(F^A_{xy}/2\Big) &= \frac{1}{2}\sum_{xy}\left(\frac{\rho(x)}{\pi(x)} - \frac{\rho(y)}{\pi(y)}\right)(q_{xy} - q_{yx})\\ &= \sum_x \frac{\rho(x)}{\pi(x)}\sum_y (q_{xy} - q_{yx}) = 0,\end{aligned}$$

where the last equality uses (7). This establishes (28). □

In Sect. 4.4, we will reformulate the so-called Hamilton–Jacobi relation of MFT in terms of forces, and show that this yields an equation analogous to (28).

3.2 Physical Interpretation of F^S and F^A

In stochastic thermodynamics, one may identify F^A_{xy} as the *housekeeping heat* (or *adiabatic entropy production*) associated with a single transition from state x to state y, see [16,50]. (Within the Markov chain formalism, there is some mixing of the notions of force and energy: usually an energy would be a product of a force and a distance but there is no notion of a distance between states of the Markov chain, so forces and energies have the same units in our analysis.) Hence $j \cdot F^A$ is the rate of flow of housekeeping heat into the environment. The meaning of the housekeeping heat is that for irreversible systems, transitions between states involve unavoidable dissipated heat which cannot be transformed into work (this dissipation is required in order to "do the housekeeping").

To obtain the physical interpretation of F^S, we also define

$$D(\rho, j) := \frac{1}{2}\sum_{xy} j_{xy}\log\frac{\rho(y)\pi(x)}{\rho(x)\pi(y)}. \tag{29}$$

For a general path $(\rho_t, j_t)_{t\in[0,T]}$ that satisfies $\dot\rho_t = -\operatorname{div} j_t$, we also identify

$$\frac{d}{dt}\mathcal{F}(\rho_t) = \sum_x \dot\rho_t(x)\log\frac{\rho_t(x)}{\pi(x)} = \frac{1}{2}\sum_{xy}(j_t)_{xy}\nabla^{x,y}\log\frac{\rho}{\pi} = D(\rho_t, j_t), \tag{30}$$

where we used (8), (15). That is, $D(\rho, j)$ is the change in free energy induced by the current j. Moreover it is easy to see that

$$F^S_{xy}(\rho) = -\nabla^{x,y}\frac{\delta\mathcal{F}}{\delta\rho}, \tag{31}$$

where $\frac{\delta\mathcal{F}}{\delta\rho}$ denotes the functional derivative of the free energy $\mathcal{F}$ given in (8). (Note that the functional derivative $\delta\mathcal{F}/\delta\rho$ is simply $\partial\mathcal{F}/\partial\rho$ in this case, since ρ is defined on a discrete space. We retain the functional notation to emphasise the connection to the general setting of Sect. 1.1.) Also, the last identity in (30) can be phrased as

$$j \cdot F^S(\rho) = -D(\rho, j). \tag{32}$$

The same identity, with an integration by parts, shows that

$$D(\rho, j) = 0 \text{ if } j \text{ is divergence free.} \tag{33}$$

Equation (31) shows that the symmetric force F^S is minus the gradient of the free energy, so the heat flow associated with the dual pairing of j and F^S is equal to (the negative of) the rate of change of the free energy. It follows that the right hand side of (25) can alternatively be written as $-\int j \cdot F^A \, \mathrm{d}t$.

We also recall from Sect. 2.2 that the force F acts in the space of probability densities: F_{xy} depends not only on the states x, y but also on the density ρ. (Physical forces acting on individual copies of the system should not depend on ρ since each copy evolves independently, but F includes entropic terms associated with the ensemble of copies.) To understand this dependence, it is useful to write $\mathcal{F}(\rho) = -\sum_x \rho(x) \log \pi(x) + \sum_x \rho(x) \log \rho(x)$. We also write the invariant measure in a Gibbs-Boltzmann form: $\pi(x) = \exp(-U(x))/Z$, where $U(x)$ is the internal energy of state x and $Z = \sum_x \exp(-U(x))$ is a normalisation constant. Then $-\sum_x \rho(x) \log \pi(x) = \mathbb{E}_\rho(U) + \log Z$ depends on the mean energy of the system, while $\sum_x \rho(x) \log \rho(x)$ is (the negative of) the mixing entropy, which comes from the many possible permutations of the copies of the system among the states of the Markov chain. From (31) one then sees that F^S has two contributions: one term (independent of ρ) that comes from the gradient of the energy U and the other (which depends on ρ) comes from the gradient of the entropy. These entropic forces account for the fact that a given empirical density $\rho^{\mathcal{N}}$ can be achieved in many different ways, since individual copies of the system can be permuted among the different states of the system.

3.3 Generalised Orthogonality for Forces

Recalling the definitions of Sect. 3.1, one sees that the current in the adjoint process satisfies an analogue of (13):

$$J^*_{xy}(\rho) := a_{xy}(\rho) \sinh\big(\tfrac{1}{2} F^*_{xy}(\rho)\big), \qquad \text{with} \qquad F^*_{xy}(\rho) := F^S_{xy}(\rho) - F^A_{xy}. \tag{34}$$

Comparing with (27), one sees that the adjoint process may also be obtained by inverting F^A (while keeping $F^S(\rho)$ as it is). For $a^S_{xy}(\rho) := a_{xy}(\rho) \cosh(F^A_{xy}/2)$ the symmetric current is defined as

$$J^S_{xy}(\rho) := a^S_{xy}(\rho) \sinh\big(F^S_{xy}(\rho)/2\big), \tag{35}$$

which satisfies $J^S_{xy}(\rho) = (J_{xy}(\rho) + J^*_{xy}(\rho))/2$. It is the same for the process and the adjoint process, and also coincides with the current for reversible processes (where $q_{xy} = q_{yx}$, or equivalently $F^A = 0$). An analogous formula can also be obtained for the anti-symmetric current. With $a^A_{xy}(\rho) := a_{xy}(\rho) \cosh(F^S_{xy}(\rho)/2) = a_{xy}(\pi)\big(\frac{\rho(x)}{\pi(x)} + \frac{\rho(y)}{\pi(y)}\big)/2$, the anti-symmetric current is defined as

$$J^A_{xy}(\rho) := a^A_{xy}(\rho) \sinh\big(F^A_{xy}/2\big). \tag{36}$$

It satisfies $J^A_{xy}(\rho) = (J_{xy}(\rho) - J^*_{xy}(\rho))/2$.

Let $\Psi^\star_S$ be the symmetric version of $\Psi^\star$ obtained from (16) with $a_{xy}(\rho)$ replaced by $a^S_{xy}(\rho)$. (The Legendre transform of $\Psi^\star_S$ is similarly denoted Ψ_S). This leads to a separation of $\Psi^\star(\rho, F(\rho))$ in a term corresponding to $F^S(\rho)$ and a term corresponding to F^A.

Lemma 2 *The two forces $F^S(\rho)$ and F^A defined in (27) satisfy*

$$\Psi^\star(\rho, F(\rho)) = \Psi^\star_S\big(\rho, F^S(\rho)\big) + \Psi^\star\big(\rho, F^A\big), \tag{37}$$

Proof Using $\cosh(x+y) = \cosh(x)\cosh(y) + \sinh(x)\sinh(y)$, Lemma 1 and the definition of $a^S_{xy}(\rho)$, we obtain that the left hand side of (37) is given by

$$\sum_{xy} a_{xy}(\rho)\big(\cosh(F_{xy}(\rho)/2) - 1\big) = \sum_{xy} a_{xy}(\rho)\big(\cosh(F^S_{xy}(\rho)/2)\cosh\big(F^A_{xy}(\rho)/2\big) - 1\big)$$
$$= \sum_{xy} a^S_{xy}(\rho)\big(\cosh\big(F^S_{xy}(\rho)/2\big) - 1\big) + \sum_{xy} a_{xy}(\rho)\big(\cosh\big(F^A_{xy}(\rho)/2\big) - 1\big), \tag{38}$$

which coincides with the right hand side of (37). □

The physical interpretation of Lemma 2 is that the strength of the force $F(\rho)$ can be written as separate contributions from $F^S(\rho)$ and F^A. The following corollary allows us to think of a generalised orthogonality of the forces $F^S(\rho)$ and F^A.

Proposition 3 (Generalised orthogonality) *The forces $F^S(\rho)$ and F^A satisfy*

$$\Psi^\star\big(\rho, F^S(\rho) + F^A\big) = \Psi^\star\big(\rho, F^S(\rho) - F^A\big). \tag{39}$$

Proof This follows directly from Lemma 2 and the symmetry of $\Psi^\star(\rho, \cdot)$. □

We refer to Proposition 3 as a generalised orthogonality between F^S and F^A because $\Psi^\star$ is acting as generalisation of a squared norm (see Sect. 1.1), so (39) can be viewed as a nonlinear generalisation of $\|F^S + F^A\|^2 = \|F^S - F^A\|^2$, which would be a standard orthogonality between forces.

Moreover, Lemma 2 can be used to decompose the OM functional as a sum of three terms.

Corollary 4 *Let Φ_S be defined as in (3) with $(\Psi, \Psi^\star)$ replaced by $(\Psi_S, \Psi_S^\star)$, and $D(\rho, j)$ as defined in (29). Then*

$$\Phi(\rho, j, F(\rho)) = D(\rho, j) + \Phi_S\big(\rho, 0, F^S(\rho)\big) + \Phi\big(\rho, j, F^A\big). \tag{40}$$

Proof We use the definition of Φ in (3) and (32) together with Lemma 2 to decompose $\Phi(\rho, j, F(\rho))$ as

$$\begin{aligned}\Phi(\rho, j, F(\rho)) &= D(\rho, j) + \Psi_S^\star\big(\rho, F^S(\rho)\big) + \Big[\Psi(\rho, j) - j \cdot F^A + \Psi^\star\big(\rho, F^A\big)\Big] \\ &= D(\rho, j) + \Phi_S\big(\rho, 0, F^S(\rho)\big) + \Phi\big(\rho, j, F^A\big),\end{aligned} \tag{41}$$

which proves the claim. □

Recall from Sect. 1.1 that Φ measures how much the current j deviates from the typical (or most likely) current $J(\rho)$. One sees from (40) that it can be large for three reasons. The first term is large if the current is pushing the system up in free energy (because D is the rate of change of free energy induced by the current j). The second term comes from the time-reversal symmetric (gradient) force $F^S(\rho)$, which is pushing the system towards equilibrium. The third term comes from the time-reversal anti-symmetric force F^A; namely, it measures how far the current j is from the value induced by the force F^A.

Corollary 4 also makes it apparent that the free energy $\mathcal{F}$ is monotonically decreasing for solutions of (11), which are minimisers of $I_{[0,T]}$.

Corollary 5 *The free energy $\mathcal{F}$ is monotonically decreasing along minimisers of the rate function $I_{[0,T]}$. Its rate of change is given by*

$$\frac{d}{dt}\mathcal{F}(\rho_t) = -\Psi_S^\star\big(\rho_t, F^S(\rho_t)\big) - \Phi\big(\rho_t, J(\rho_t), F^A(\rho_t)\big). \tag{42}$$

Proof For minimisers of the rate function one has $\Phi = 0$. Hence (30) and Corollary 4 imply that

$$\frac{d}{dt}\mathcal{F}(\rho_t) = D(\rho, j) = -\Psi_S^\star\big(\rho_t, F^S(\rho_t)\big) - \Phi\big(\rho_t, J(\rho_t), F^A(\rho_t)\big). \tag{43}$$

Both $\Psi^\star$ and Φ are non-negative, so $\mathcal{F}$ is indeed monotonically decreasing. □

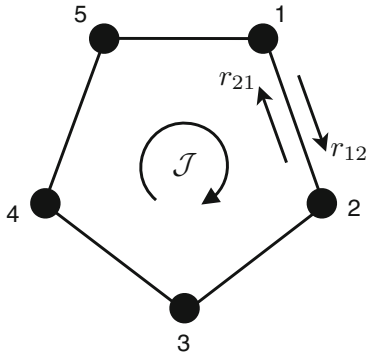

Fig. 1 Illustration of a simple Markov chain with $n = 5$ states arranged in a circle. The transition rates between states are $r_{i,i\pm1}$. If the Markov chain is not reversible, there will be a steady-state probability current $\mathcal{J}$ corresponding to a net drift of the system around the circle

3.4 Hamilton–Jacobi Like Equation for Markov Chains

It is also useful to note at this point an additional aspect of the orthogonality relationships presented here, which has connections to MFT (see Sect. 4). We formulate an analogue of the Hamilton–Jacobi equation of MFT, as follows. Define

$$\mathbb{H}(\rho, \xi) = \frac{1}{2}\left[\Psi^\star(\rho, F(\rho) + 2\xi) - \Psi^\star(\rho, F(\rho))\right], \tag{44}$$

which we refer to as an *extended Hamiltonian*, for reasons discussed in Sect. 6.3 (see also Sect. IV.G of [7]).

The *extended Hamilton–Jacobi equation* for a functional $\mathcal{S}$ is then (cf. equation (100) in Sect. 6.3) given by

$$\mathbb{H}\left(\rho, \nabla\frac{\delta\mathcal{S}}{\delta\rho}\right) = 0. \tag{45}$$

Note that the free energy $\mathcal{F}$ defined in (8) solves (45), which follows from Proposition 3 (using (31) and that $\Psi^\star$ is symmetric in its second argument). In fact (see Proposition 13), the free energy is the maximal solution to this equation. In MFT, the analogous variational principle can be useful, as a characterisation of the invariant measure of the process. Here, one has a similar characterisation of the (non-equilibrium) free energy.

Since (45) with $\mathcal{S} = \mathcal{F}$ provides a characterisation of the free energy $\mathcal{F}$, which is uniquely determined by the invariant measure π of the process, it follows that (45) must be equivalent to the condition that π satisfies div $J(\pi) = 0$: recall (11). Writing everything in terms of the rates of the Markov chain and its adjoint, (45) becomes

$$\sum_x \rho(x) \sum_y [r_{xy} - r^*_{xy}] = 0,$$

which must hold for all ρ: from the definition of r^* one then has $\sum_y \pi(x) r_{xy} = \sum_y \pi(y) r_{yx}$, which is indeed satisfied if and only if π is invariant (cf. Eq. (7)).

3.5 Example: Simple Ring Network

To illustrate these abstract ideas, we consider a very simple Markov chain, in which n states are arranged in a circle, see Fig. 1. So $V = \{1, 2, \ldots, n\}$ and the only allowed transitions take place between state x and states $x \pm 1$ (to incorporate the circular geometry we interpret $n+1 = 1$ and $1-1 = n$). In physics, such Markov chains arise (for example) as simple models of nano-machines or motors, where an external energy source might be used to drive circular motion [29,53]. Alternatively, such a Markov chain might describe a protein molecule that

goes through a cyclic sequence of conformations, as it catalyses a chemical reaction [31]. In both cases, the systems evolve stochastically because the relevant objects have sizes on the nano-scale, so thermal fluctuations play an important role.

To apply the analysis presented here, the first step is to identify forces and mobilities, as in (12). Let $R_x = \sqrt{r_{x,x+1}r_{x+1,x}}$. The invariant measure may be identified by solving $\sum_y \pi(x)r_{xy} = \sum_y \pi(y)r_{yx}$ subject to $\sum_y \pi(y) = 1$. Finally, one computes the steady state current $\mathcal{J} = \pi(x)r_{x,x+1} - \pi(x+1)r_{x+1,x}$, where the right hand side is independent of x (this follows from the steady-state condition on π). The original Markov process has $2n$ parameters, which are the rates $r_{x,x\pm1}$: these are completely determined by the $n-1$ independent elements of π, the n mobilities $(R_x)_{x=1}^n$ and the current $\mathcal{J}$. The idea is that this reparameterisation allows access to the physically important quantities in the system.

From the definitions of $\mathcal{J}$ and R, it may be verified that

$$2\pi(x)r_{x,x+1} = \sqrt{\mathcal{J}^2 + 4R_x^2\pi(x)\pi(x+1)} + \mathcal{J},$$

and similarly $2\pi(x+1)r_{x+1,x} = \sqrt{\mathcal{J}^2 + 4R_x^2\pi(x)\pi(x+1)} - \mathcal{J}$. Then write

$$\rho(x)r_{x,x+1} = R_x\sqrt{\rho(x)\rho(x+1)} \times \sqrt{\frac{\rho(x)\pi(x+1)}{\rho(x+1)\pi(x)}} \cdot \times \left(\frac{\sqrt{\mathcal{J}^2 + 4R_x^2\pi(x)\pi(x+1)} + \mathcal{J}}{\sqrt{\mathcal{J}^2 + 4R_x^2\pi(x)\pi(x+1)} - \mathcal{J}}\right)^{1/2}. \tag{46}$$

In this case, we can identify the three terms as

$$\rho(x)r_{x,x+1} = \frac{1}{2}a_{x,x+1}(\rho) \times \exp\left(F^S_{x,x+1}(\rho)/2\right) \times \exp\left(F^A_{x,x+1}/2\right), \tag{47}$$

which allows us to read off the mobility a and the forces F^S and F^A. The physical meaning of these quantities may not be obvious from these definitions, but we show in the following that reparameterising the transition rates in this way reveals structure in the dynamical fluctuations.

For example, equilibrium models (with detailed balance) can be identified via $F^A_{x,x+1} = 0$ (for all x). In general $F^A_{x,x+1}$ is the (steady-state) entropy production associated with a transition from x to $x+1$, see Sect. 3.2. The steady state entropy production associated with going once round the circuit is $\sum_x F^A_{x,x+1} = \log\prod_x(r_{x,x+1}/r_{x+1,x})$, as it must be [1].

Now consider the LDP in (23). We consider a large number ($\mathcal{N}$) of identical nano-scale devices, each of which is described by an independent copy of the Markov chain. Typically, each device goes around the circle at random, and the average current is $\mathcal{J}$ (so each object performs $\mathcal{J}/n$ cycles per unit time). The LDP describes properties of the ensemble of devices. If $\mathcal{N}$ is large and the distribution of devices over states is ρ, then the (overwhelmingly likely) time evolution of this distribution is $\dot\rho = -\operatorname{div} J(\rho)$, where the current J obeys the simple formula

$$J_{x,x+1}(\rho) = a_{x,x+1}(\rho)\sinh\left(\tfrac{1}{2}\left[F^S_{x,x+1}(\rho) + F^A_{x,x+1}\right]\right), \tag{48}$$

which is (13), applied to this system. The simplicity of this expression motivates the parametrisation of the transition rates in terms of forces and mobilities. In addition, if one observes some current j [not necessarily equal to $J(\rho)$] then the rate of change of free energy of the ensemble can be written compactly as $D(\rho, j) = -j\cdot F^S(\rho)$, from (32). The quantity $j\cdot F^A$ is the rate of dissipation via housekeeping heat (see Sect. 3.2). This (physically-motivated)

splitting of $j \cdot F = j \cdot (F^S + F^A)$ motivates our introduction of the two forces F^S and F^A. Note that $j \cdot F$ is the rate of heat flow from the system to its environment, and appears in the fluctuation theorem (25).

Finally we turn to the large deviations of this ensemble of nano-scale objects. There is an LDP (23), whose rate function can be decomposed into three pieces (Corollary 4), because of the generalised orthogonality of the forces F^S and F^A (Lemma 2). This splitting of the rate function is useful because the symmetry properties of the various terms yields bounds on rate functions for some other LDPs obtained from Φ by contraction, see Sect. 5.

4 Connections to MFT

MFT is a field theory which describes the mass evolution of particle systems in the drift-diffusive regime, on the level of hydrodynamics. In this setting, it can be seen as generalisation of Onsager–Machlup theory [36]. For a comprehensive review, we refer to [7]. This section gives an overview of the theory, focussing on the connections to the results presented in Sects. 2 and 3.

We seek to emphasise two points: first, while the particle currents in MFT and the probability current in Markov chains are very different objects, they both obey large-deviation principles of the form presented in Sect. 1.1. This illustrates the broad applicability of this general setting. Second, we note that many of the particle models for which MFT gives a macroscopic description are Markov chains on discrete spaces. Starting from this observation, we argue in Sect. 4.5 that some results that are well-known in MFT originate from properties of these underlying Markov chains, particularly Proposition 3 and Corollary 4.

4.1 Setting

We consider a large number N of indistinguishable particles, moving on a lattice Λ_L (indexed by $L \in \mathbb{N}$, such that the number of sites $|\Lambda_L|$ is strictly increasing with L). These particles are described by a Markov chain, so the relevant forces and currents satisfy the equations derived in Sects. 2 and 3. The hydrodynamic limit is obtained by letting $L \to \infty$ such that the total density $N/|\Lambda_L|$ converges to a fixed number $\bar{\rho}$. In this limit, the lattice Λ_L is rescaled into a domain $\Lambda \subset \mathbb{R}^d$ and one can characterise the system by a local (mass) density $\rho \colon \Lambda \to [0, \infty)$ together with a local current $j \colon \Lambda \to \mathbb{R}^d$, which evolve deterministically as a function of time [7,28]. This time evolution depends on some (density-dependent) applied forces $F(\rho) \colon \Lambda \to \mathbb{R}^d$. The force at $x \in \Lambda$ can be written as

$$F(\rho)(x) = \hat{f}''(\rho(x))\nabla\rho(x) + E(x), \tag{49}$$

where the gradient ∇ denotes a spatial derivative, the function $\hat{f} \colon [0, \infty) \to \mathbb{R}$ is a free energy density and $E \colon \Lambda \to \mathbb{R}^d$ is a drift. (The free energy $\hat{f}$ is conventionally denoted by f [7]; here we use a different notation since f indicates a force in this work.) With these definitions, the deterministic currents satisfy the linear relation [41]

$$J(\rho) = \chi(\rho)F(\rho), \tag{50}$$

which is the hydrodynamic analogue of (13). Here, $\chi(\rho) \in \mathbb{R}^{d\times d}$ is a (density-dependent) mobility matrix.

4.2 Onsager–Machlup Functional

Within MFT, the system is fully specified once the functions f, χ, E are given. These three quantities are sufficient to specify both the deterministic evolution of the most likely path ρ, and the fluctuations away from it. We can again define an OM functional given by

$$\Phi_{\mathrm{MFT}}(\rho, j, f) := \frac{1}{2}\int_{\Lambda} \big(j - \chi f\big) \cdot \chi^{-1}\big(j - \chi f\big)\,\mathrm{d}x. \tag{51}$$

To cast this functional in the form (3), we define the dual pair $\int_{\Lambda}(j \cdot f)\,\mathrm{d}x$, together with the Legendre duals

$$\Psi_{\mathrm{MFT}}(\rho, j) := \frac{1}{2}\int_{\Lambda} j \cdot \chi^{-1} j\,\mathrm{d}x \quad \text{and} \quad \Psi^{\star}_{\mathrm{MFT}}(\rho, f) := \frac{1}{2}\int_{\Lambda} f \cdot \chi f\,\mathrm{d}x. \tag{52}$$

Given ρ and f, we have that Φ_{MFT} is uniquely minimised (and equal to zero) for the current $j = \chi(\rho) f$.

4.3 Large Deviation Principle

Within MFT, one considers an empirical density and an empirical current. We emphasise that these refer to particles, which are interacting and move on the lattice Λ_L; this is in contrast to the case of Markov chains, where the copies of the system were non-interacting and one considers a density and current of probability. The averaged number of particles at site $i \in \Lambda_L$ is denoted with $\hat{\rho}^L_t(x_i)$, where x_i is the image in the rescaled domain Λ of site $i \in \Lambda_L$, and the corresponding particle current is given by $\hat{\jmath}^L_t$ (cf. Sect. VIII.F in [7] for details). Note that both the particle density $\hat{\rho}^L_t$ and the particle current $\hat{\jmath}^L_t$ are random quantities (see also Sect. 4.5).

In keeping with the setting of Sect. 1.1, we focus on paths $(\hat{\rho}^L_t, \hat{\jmath}^L_t)_{t\in[0,T]}$ in the limit as $L \to \infty$, where the probability is, analogous to (1), given by

$$\mathrm{Prob}\Big(\big(\hat{\rho}^L_t, \hat{\jmath}^L_t\big)_{t\in[0,T]} \approx (\rho_t, j_t)_{t\in[0,T]}\Big) \asymp \exp\big\{-|\Lambda_L| I^{\mathrm{MFT}}_{[0,T]}\big((\rho_t, j_t)_{t\in[0,T]}\big)\big\}. \tag{53}$$

Note that the parameter $\mathcal{N}$ in (1), which is the speed of the LDP, corresponds to the lattice size $|\Lambda_L|$. For the force $F(\rho)$ defined in (49), the rate functional in (53) is given by

$$I^{\mathrm{MFT}}_{[0,T]}\big((\rho_t, j_t)_{t\in[0,T]}\big) = \begin{cases} \mathcal{V}(\rho_0) + \frac{1}{2}\int_0^T \Phi_{\mathrm{MFT}}(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t & \text{if } \dot{\rho}_t + \mathrm{div}\, j_t = 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{54}$$

Here $\mathcal{V}$ is the *quasipotential*, which plays the role of a non-equilibrium free energy. We may think of $\mathcal{V}$ as the macroscopic analogue of the free energy $\mathcal{F}$ defined in (8). It is the rate functional for the process sampled from the invariant measure, which is consistent with the case for Markov chains in (24). We assume that $\mathcal{V}$ has a unique minimiser π, which is the steady-state density profile (so $\mathcal{V}(\pi) = 0$).

An important difference between the Markov chain setting and MFT is that the OM functional for Markov chains is non-quadratic, which is equivalent to a non-linear flux force relation, whereas MFT is restricted to quadratic OM functionals.

Equation (53) is the basic assumption in MFT [7], in the sense that all systems considered by MFT are assumed to satisfy this pathwise LDP. In fact, both the process and its adjoint are assumed to satisfy such LDPs (with similar rate functionals, but different forces) [7].

4.4 Decomposition of the Force F

The force F in (49) can be written as the sum of a symmetric and an anti-symmetric part, $F(\rho) = F_S(\rho) + F_A(\rho)$, just as in Sect. 3.1. The force for the adjoint process is given by $F^*(\rho) = F_S(\rho) - F_A(\rho)$. Note that, unlike in the case of Markov chains, $F_A(\rho)$ can here depend on ρ. More precisely, $F_S(\rho) = -\nabla \frac{\delta \mathcal{V}}{\delta \rho}$ and $F_A(\rho)$ is given implicitly by $F_A(\rho) = F(\rho) - F_S(\rho)$.

The symmetric and anti-symmetric currents are defined in terms of the forces $F_S(\rho)$ and $F_A(\rho)$ as $J_S(\rho) := \chi(\rho) F_S(\rho)$ and $J_A(\rho) := \chi(\rho) F_A(\rho)$. An important result in MFT is the so-called *Hamilton–Jacobi orthogonality*, which states that

$$\int_\Lambda J_S(\rho) \cdot \chi(\rho)^{-1} J_A(\rho)\, \mathrm{d}x = 0. \tag{55}$$

In terms of the forces $F_S(\rho)$ and $F_A(\rho)$, we can restate (55) as

$$\int_\Lambda F_S(\rho) \cdot \chi(\rho) F_A(\rho)\, \mathrm{d}x = 0. \tag{56}$$

The latter is the quadratic version of the orthogonality (28) of Lemma 1; it is equivalent to

$$\begin{aligned} &\int_\Lambda \big(F_S(\rho) + F_A(\rho)\big) \cdot \chi(\rho) \big(F_S(\rho) + F_A(\rho)\big)\, \mathrm{d}x \\ &\quad = \int_\Lambda \big(F_S(\rho) - F_A(\rho)\big) \cdot \chi(\rho) \big(F_S(\rho) - F_A(\rho)\big)\, \mathrm{d}x, \end{aligned} \tag{57}$$

or in other words, from (52),

$$\Psi^\star_{\mathrm{MFT}}(\rho, F_S(\rho) + F_A(\rho)) = \Psi^\star_{\mathrm{MFT}}(\rho, F_S(\rho) - F_A(\rho)), \tag{58}$$

which is the result of Proposition 3 in the context of MFT. One can see (39), and hence Proposition 3, as the natural generalisation to the Hamilton–Jacobi orthogonality (55). Again, the MFT describes systems on the macroscopic scale, but the result (58) originates from the result (39), on the microscopic level.

4.5 Relating Markov Chains to MFT: Hydrodynamic Limits

We have discussed a formal analogy between current/density fluctuations in Markov chains and in MFT: the large deviation principles (23) and (53) refer to different objects and different limits, but they both fall within the general setting described in Sect. 1.1. We argue here that the similarities between these two large deviation principles are not coincidental—they arise naturally when MFT is interpreted as a theory for hydrodynamic limits of interacting particle systems.

To avoid confusion between particle densities and probability densities, we introduce (only for this section) a different notation for some properties of discrete Markov chains, which is standard for interacting particle systems. Let η represent a state of the Markov chain (in place of the notation x of Sect. 2), and let μ be a probability distribution over these states (in place of the notation ρ of Sect. 2). Let $\jmath$ be the probability current.

We illustrate our argument using the weakly asymmetric simple exclusion process (WASEP) in one dimension, so the lattice is $\Lambda_L = \{1, 2, \ldots, L\}$, and each lattice site contains at most one particle, so $V = \{0, 1\}^L$. The lattice has periodic boundary conditions and the occupancy of site i is $\eta(i)$. Particles hop to the right with rate L^2 and to the left with rate $L^2(1 - (E/L))$, but in either case only if the destination site is empty. Here E is a fixed

parameter (an external field); the dependence of the hop rates on L is chosen to ensure a diffusive hydrodynamic limit (as required for MFT).

The spatial domain relevant for MFT is $\Lambda = [0, 1]$: site $i \in \Lambda_L$ corresponds to position $i/L \in \Lambda$. For any probability measure μ on V, one can write a corresponding smoothed particle density ρ^ϵ on Λ, as

$$\rho^\epsilon(x) = \frac{1}{L} \sum_{\eta \in V} \sum_{i=1}^{L} \mu(\eta)\eta(i/L)\delta^\epsilon(x - (i/L)), \tag{59}$$

where δ^ϵ is a smoothed delta function (for example a Gaussian with unit weight and width ϵ, or—more classically—a top-hat function of width ϵ, cf. [28]). Similarly if there is a probability current $\jmath$ in the Markov chain, one can write a smoothed particle current as

$$j^\epsilon(x) = \frac{1}{L} \sum_{\eta \in V} \sum_{i=1}^{L} \jmath_{\eta,\eta^{i,i+1}} \delta^\epsilon \left(x - \frac{2i+1}{2L} \right), \tag{60}$$

where $\eta^{i,i+1}$ is the configuration obtained from η by moving a particle from site i to site $i + 1$; if there is no particle on site i then define $\eta^{i,i+1} = \eta$ so that $\jmath_{\eta,\eta^{i,i+1}} = 0$. Physically, ρ^ϵ is the average particle density associated to μ, and j^ϵ is the particle current associated to $\jmath$.

As noted above, MFT is concerned with the limit $L \to \infty$. The LDP (23) is not relevant for that limit (it applies when one considers many ($\mathcal{N} \to \infty$) independent copies of the Markov chain, with L being finite for each copy). However, the rate function $I_{[0,T]}$ that appears in (23) has an alternative physical interpretation, as the relative entropy between two path measures: see Appendix A. This relative entropy can be seen as a property of the WASEP; there is no requirement to invoke many copies of the system. Physically, the relative entropy measures how different is the WASEP from an alternative Markov process with a given probability and current $(\mu_t, \jmath_t)_{t\in[0,T]}$.

The key point is that in cases where MFT applies, one expects that the rate function $I^{\mathrm{MFT}}_{[0,T]}$ can be related to this relative entropy. In fact, there is a deeper relation between relative entropies and rate functionals: it can be shown that Large Deviation Principles are equivalent to Γ-convergence of relative entropy functionals (see [42] for details).

Returning to the WASEP, we consider a particle density $(\rho_t, j_t)_{t\in[0,T]}$ that satisfies $\dot\rho_t = -\operatorname{div} j_t$. One then can find (for each L) a time-dependent probability and current $(\mu^L_t, \jmath^L_t)_{t\in[0,T]}$, with $\dot\mu^L_t = -\operatorname{div} \jmath^L_t$, such on taking the limit $\epsilon \to 0$ *after* $L \to \infty$, the associated particle densities $(\rho^\epsilon_t, j^\epsilon_t) \to (\rho_t, j_t)$ and moreover

$$\lim_{L\to\infty} \frac{1}{|\Lambda_L|} I_{[0,T]}\big((\mu^L_t, \jmath^L_t)_{t\in[0,T]}\big) = I^{\mathrm{MFT}}_{[0,T]}\big((\rho_t, j_t)_{t\in[0,T]}\big). \tag{61}$$

In order to find $(\mu^L_t, \jmath^L_t)_{t\in[0,T]}$, one defines a "controlled" WASEP (similar to (69) in Sect. 5.3), in which the particle hop rates depend on position and time, such that the particle density in the hydrodynamic limit obeys $\dot\rho_t = -\operatorname{div} j_t$.

For interacting particle systems, this "controlled" process is usually obtained by adding a time dependent external field to the system that acts on the individual particles. This was first derived for the symmetric SEP in [27] (see also [4] for a treatment of the zero-range process). For the WASEP (in a slightly different situation with open boundaries) a proof of (61) can e.g. be found in [6], Lemma 3.7.

Moreover, on decomposing $I^{\mathrm{MFT}}_{[0,T]}$ and $I_{[0,T]}$ as in (3), the separate functions Ψ and $\Psi^\star$ obey formulae analogous to (61): this is the sense in which the structure of the MFT rate

function is inherited from the relative entropy of the Markov chains. The quadratic functions Ψ and $\Psi^\star$ in MFT arise because the forces that appear in the underlying Markov chains are small (compared to unity), so second order Taylor expansions of $\Psi^\star$ and Ψ give in the limit the accurate description, similar to [2]. We will return to this discussion in a later publication.

5 LDPs for Time-Averaged Quantities

So far we have considered large deviation principles for hydrodynamic limits, and for systems consisting of many independent copies of a single Markov chain. We now show how some of the results derived in Sects. 2 and 3 also have analogues for large deviations for a single Markov chain, in the large-time limit.

5.1 Large Deviations at Level 2.5

Analogous to (22), we define the time averaged empirical measure of a single copy of the Markov chain $\hat{\rho}_{[0,T]}$ and the time averaged empirical current $\hat{j}_{[0,T]}$ as

$$\hat{\rho}_{[0,T]} := \frac{1}{T}\int_0^T \hat{\rho}_t \,\mathrm{d}t \quad \text{and} \quad \hat{j}_{[0,T]} := \frac{1}{T}\int_0^T \hat{j}_t \,\mathrm{d}t \tag{62}$$

(where we choose $\hat{\rho}_t = \hat{\rho}_t^1$ and $\hat{j}_t = \hat{j}_t^1$ for the empirical density and current of the single Markov chain, as defined above in Sect. 2.3). For countable state Markov chains, the quantity $(\hat{\rho}_{[0,T]}, \hat{j}_{[0,T]})$ satisfies a LDP as $T \to \infty$:

$$\mathrm{Prob}\big((\hat{\rho}_{[0,T]}, \hat{j}_{[0,T]}) \approx (\rho, j)\big) \asymp \exp\big\{-T I_{2.5}(\rho, j)\big\}. \tag{63}$$

We refer to such principles as *level 2.5 LDPs*. For countable state Markov chains the rate functional $I_{2.5}(\rho, j)$ was derived in [39], and was proven rigorously in [8,9] for Markov chains in the setting of Sect. 2.1 under some additional conditions (see [8,9] for the details). We can recast the rate functional (see [8, Theorem 6.1]) as

$$I_{2.5}(\rho, j) = \begin{cases} \frac{1}{2}\Phi(\rho, j, F(\rho)) & \text{if } \operatorname{div} j = 0 \\ +\infty & \text{otherwise} \end{cases}, \tag{64}$$

with Φ again given by (3), together with (14), (16) and (18).

We have stated this LDP for joint fluctuations of the density and the current. For Markov chains, the LDP for the density and the *flow* is also known as a level-2.5 LDP [9], so our general use of the name level-2.5 for (63) may be non-standard, but it seems reasonable. The rate functional for the density and the current in (63) can be obtained by contraction from the rate functional for the density and the flow (see Theorem 6.1 in [8]).

Using the splitting obtained in Sect. 3.3, we obtain the following representation for the rate functional on level-2.5.

Proposition 6 *Let j be divergence free. Then the level-2.5 rate functional (64) is given by*

$$I_{2.5}(\rho, j) = \frac{1}{2}\Big[\Phi_S\big(\rho, 0, F^S(\rho)\big) + \Phi\big(\rho, j, F^A\big)\Big]. \tag{65}$$

Proof We note from (33) that $D(\rho, j)$ vanishes for divergence free currents j. The result then directly follows from Corollary 4. □

5.2 Large Deviations for Currents

Proposition 6 is connected to recently-derived bounds on rate functions for currents, see [22, 23,45,46]. Indeed, the rate function for current fluctuations can be obtained by contraction from level-2.5, as

$$I_{\text{current}}(j) := \inf_{\rho} I_{2.5}(\rho, j). \tag{66}$$

Then, following [23,46], it may be shown that for any ρ, j, f one has for Φ as in (3) with (14), (16)–(18) that

$$\Phi\big(\rho, j, f\big) \leq \sum_{xy} \Big(j_{xy} - j^f_{xy}(\rho)\Big)^2 b_{xy}(\rho, f) \tag{67}$$

with $b_{xy}(\rho, f) = f_{xy}/(4j^f_{xy}(\rho))$ if $f_{xy} \neq 0$; otherwise b_{xy} is continuously extended by taking $b_{xy}(\rho, f) = 1/(2a_{xy}(\rho))$. Hence one has the result of [22], that the curvature of the rate function is controlled by the housekeeping heat F^A, as

$$I_{\text{current}}(j) \leq I_{2.5}(\pi, j) = \frac{1}{2}\Phi\big(\pi, j, F^A\big) \leq \frac{1}{2}\sum_{xy} \frac{\big(j_{xy} - J^{\text{ss}}_{xy}\big)^2}{4(J^{\text{ss}}_{xy})^2} J^{\text{ss}}_{xy} F^A_{xy}, \tag{68}$$

where $J^{\text{ss}} := J(\pi)$ is the steady state current (recall (9)), and the ratio $F^A_{xy}/J^{\text{ss}}_{xy}$ must again be interpreted as $2/a_{xy}(\rho)$ in the case where F^A_{xy} (and hence J^{ss}_{xy}) vanish. The first step in (68) comes from (66), the second step uses (65) as well as $\Phi(\pi, 0, F^S) = 0$, and the third uses (67).

The significance of the splitting (65) for this result is that $J^{\text{ss}}_{xy} F^A_{xy}$ is the rate of flow of housekeeping heat associated with edge xy: the appearance of the housekeeping heat is natural since the bound comes from the second term in (65), which is independent of F^S and depends only on F^A.

5.3 Optimal Control Theory

It will be useful to introduce ideas of optimal control theory, whose relationship with large deviation theory is discussed in [10,11,18,24]. In parallel with our given transition rates r_{xy} we introduce a new process, the *controlled process*, where the rates are modified by a *control potential* φ, as

$$\tilde{r}_{xy} := r_{xy} \exp((\varphi(y) - \varphi(x))/2). \tag{69}$$

For a given probability distribution ρ, we seek a potential φ such that the controlled process has invariant measure $\tilde{\pi} := \rho$. For this we need

$$\sum_y \big[\rho_x r_{xy} \exp((\varphi(y) - \varphi(x))/2) - \rho_y r_{yx} \exp((\varphi(x) - \varphi(y))/2)\big] = 0,$$

or equivalently

$$\text{div}\, j^{F+\nabla\varphi}(\rho) = \sum_y a_{xy}(\rho) \sinh\big((F_{xy}(\rho) + \nabla^{x,y}\varphi)/2\big) = 0. \tag{70}$$

We stress that, for any fixed ρ, (70) is equivalent to solving the minimisation problem

$$\inf_{\text{div}\, j = 0} \Phi\big(\rho, j, F(\rho)\big), \tag{71}$$

which is also equivalent to maximisation of the Donsker–Varadhan functional, see for example Chapter IV.4 in [15]. A proof for the existence and uniqueness of φ can, e.g., be found in [40]. Now assume that φ solves (70). The resulting controlled process depends on ρ and has rates $\tilde{r}$ given by (69). Throughout this section, we use tildes to indicate properties of the controlled process: all these quantities depend implicitly on the fixed probability ρ. Hence the (time-dependent) measure of the controlled process is $\tilde{\rho}$.

Repeating the analysis of Sect. 2.1 and noting that $\tilde{r}_{xy}\tilde{r}_{yx} = r_{xy}r_{yx}$, we find that $\tilde{a}_{xy}(\tilde{\rho}) := 2\sqrt{\tilde{\rho}(x)\tilde{r}_{xy}\tilde{\rho}(y)\tilde{r}_{yx}} = a_{xy}(\tilde{\rho})$. Also, the force for the controlled process is

$$\tilde{F}(\tilde{\rho}) = F(\tilde{\rho}) + \nabla\varphi, \tag{72}$$

which may be decomposed as

$$\begin{aligned} \tilde{F}^S(\tilde{\rho}) &:= F^S(\tilde{\rho}) + \nabla \log \frac{\rho}{\pi} = -\nabla \log \frac{\tilde{\rho}}{\rho}, \\ \tilde{F}^A &:= F(\rho) + \nabla\varphi = F^A - \nabla \log \frac{\rho}{\pi} + \nabla\varphi. \end{aligned} \tag{73}$$

Thus, the symmetric force in the controlled process vanishes when $\tilde{\rho} = \rho$. The antisymmetric force $\tilde{F}^A$ represents the force observed in the new non-equilibrium steady state ρ. If the original process is reversible, then $\varphi = \log \frac{\rho}{\pi}$ so $\tilde{F}^A = F^A = 0$.

It is useful to define $\tilde{J}_{xy}(\tilde{\rho}) := a_{xy}(\tilde{\rho}) \sinh(\tilde{F}_{xy}(\tilde{\rho})/2)$ and to identify the steady-state current for the controlled process as

$$\tilde{J}^{\mathrm{ss}} := \tilde{J}(\rho). \tag{74}$$

5.4 Decomposition of Rate Functions

The ideas of optimal control theory are useful since they facilitate the further decomposition of the level-2.5 rate function into several contributions.

Lemma 7 *Suppose that ρ and j are given and that* div $j = 0$. *Then*

$$I_{2.5}(\rho, j) = \frac{1}{2}\Big[\Phi\big(\rho, \tilde{J}^{\mathrm{ss}}, F(\rho)\big) + \Phi\big(\rho, j, \tilde{F}^A\big)\Big], \tag{75}$$

where $\tilde{J}^{\mathrm{ss}}$ is given by (74), evaluated in the optimally controlled process whose steady state is ρ.

Proof We write

$$\begin{aligned} 2I_{2.5}(\rho, j) &= \Psi(\rho, j) - j \cdot F(\rho) + \Psi^\star\big(\rho, F(\rho)\big) \\ &= [\Psi(\rho, j) - j \cdot \tilde{F}(\rho) + \Psi^\star(\rho, \tilde{F}(\rho))] \\ &\quad + \Psi^\star(\rho, F(\rho)) - \Psi^\star(\rho, \tilde{F}(\rho)) - j \cdot (F(\rho) - \tilde{F}(\rho)) \\ &= \Phi\big(\rho, j, \tilde{F}(\rho)\big) + \Psi^\star(\rho, F(\rho)) - \Psi^\star(\rho, \tilde{F}(\rho)) + j \cdot \nabla\varphi \end{aligned} \tag{76}$$

where the first line is (3) and (64); the second line is simple rewriting; and the third uses the definition of Φ in (3) and also (72) with $\tilde{\rho} = \rho$.

The current $\tilde{J}(\rho)$ satisfies $\Phi(\rho, \tilde{J}(\rho), \tilde{F}(\rho)) = 0$ so one has (by definition of Φ) that $\Psi^\star(\rho, \tilde{F}(\rho)) = \tilde{J}(\rho) \cdot \tilde{F}(\rho) - \Psi(\rho, \tilde{J}(\rho))$. Using this relation together with (72) and (76), one has

$$\begin{aligned} 2I_{2.5}(\rho, j) = \Phi\big(\rho, j, \tilde{F}(\rho)\big) + \Psi^\star(\rho, F(\rho)) - \tilde{J}(\rho) \cdot F(\rho) \\ + \Psi(\rho, \tilde{J}(\rho)) - \tilde{J}(\rho) \cdot \nabla\varphi + j \cdot \nabla\varphi. \end{aligned} \tag{77}$$

Finally we note that div $\tilde{J}(\rho) = 0$ (since ρ is the invariant measure for the controlled process) and div $j = 0$ (by assumption), so integration by parts yields $\tilde{J}(\rho) \cdot \nabla\varphi = 0 = j \cdot \nabla\varphi$; using once more the definition of Φ yields (82). □

The physical interpretation of (75) is as follows. The contribution $\frac{1}{2}\Phi(\rho, j, \tilde{F}^A)$ is a rate functional for observing an empirical current j in the controlled process, while $\frac{1}{2}\Phi(\rho, \tilde{J}^{\mathrm{ss}}, F(\rho))$ is the rate functional for observing an empirical current $\tilde{J}^{\mathrm{ss}}$ in the original process. Since $\tilde{J}^{\mathrm{ss}}$ is the (deterministic) probability current for the controlled process, one has that the more the controlled process differs from the original one, the larger will be $\Phi(\rho, \tilde{J}^{\mathrm{ss}}, F(\rho))$. Hence the level-2.5 rate functional is large if the controlled process is very different from the original one, as one might expect. The rate functional also takes larger values if the empirical current j is very different from the probability current of the controlled process.

We obtain our final representation for the level-2.5 rate functional, consisting of the sum of three different OM functionals.

Proposition 8 *Let j be divergence free. We can represent the level-2.5 rate functional* (64) *as*

$$I_{2.5}(\rho, j) = \frac{1}{2}\Big[\Phi_S\big(\rho, 0, F^S(\rho)\big) + \Phi\big(\rho, \tilde{J}^{\mathrm{ss}}, F^A\big) + \Phi\big(\rho, j, \tilde{F}^A\big)\Big]. \tag{78}$$

Proof This follows immediately from Lemma 7 followed by an application of Corollary 4 to $\Phi\big(\rho, \tilde{J}^{\mathrm{ss}}, F^A\big)$ and that $D = 0$, from (33). □

The three terms in (78) also appear in Lemma 7 and Corollary 4, and their interpretations have been discussed in the context of those results. Briefly, we recall that $I_{2.5}(\rho, j)$ sets the probability of fluctuations in which a non-typical density ρ and current j are sustained over a long time period. The first term in (78) reflects the fact that the free-energy gradient $F^S(\rho)$ tends to push ρ towards the steady state π, so maintaining any non-typical density is unlikely if $F^S(\rho)$ is large. Similarly, the second term in (78) reflects the fact that large non-gradient forces F^A also tend to suppress the probability that ρ maintains its non-typical value. The final term is the only place in which the (divergence-free) current j appears: it vanishes if the current j is typical within the controlled process (see Corollary 9); otherwise it reflects the probability cost of maintaining a non-typical circulating current.

5.5 Large Deviations at Level 2

As well the LDP (63), we also consider an (apparently) simpler object, called a *level-2 LDP*, where one considers the density only. It is formally given by

$$\mathrm{Prob}\left(\hat{\rho}_T \approx \rho\right) \asymp \exp(-T I_2(\rho)). \tag{79}$$

The contraction principle for LDPs [52, Sect. 3.6] states that

$$I_2(\rho) = \inf_{j\,:\,\mathrm{div}\, j = 0} I_{2.5}(\rho, j). \tag{80}$$

Equation (75) is uniquely minimised in its second argument for the divergence free current $j^{\tilde{F}^A}$, such that the contraction over all divergence-free vector fields j yields the level-2 rate functional

$$I_2(\rho) = \frac{1}{2}\Phi\big(\rho, \tilde{J}^{\mathrm{ss}}, F(\rho)\big). \tag{81}$$

The same splitting as above finally allows us to write the level 2 rate functional as follows.

Corollary 9 *The level-2 rate functional can be written as the sum*

$$I_2(\rho) = \frac{1}{2}\Big[\Phi_S\big(\rho, 0, F^S(\rho)\big) + \Phi\big(\rho, \tilde{J}^{\mathrm{ss}}, F^A\big)\Big]. \tag{82}$$

Proof This follows from (80) and (78), since $\Phi\big(\rho, j, \tilde{F}^A\big)$ has a minimal value of zero. □

This last identity extends the results obtained in [26] on the accelerated convergence to equilibrium for irreversible processes using LDPs from the macroscopic scale (i.e. in the regime of MFT) to Markov chains. The level-2 rate function in (82) can be interpreted as a rate of convergence to the steady state. It was shown in [26] that the rate is higher for irreversible processes, as opposed to reversible ones (as the second term $\Phi(\rho, \tilde{J}^{\mathrm{ss}}, F^A) = 0$ for reversible processes). We remark that splitting techniques for irreversible jump processes have been used to devise efficient MCMC samplers; see for example [5,34].

5.6 Connection to MFT

Under the assumption that no dynamical phase transition takes place, the time averaged density $\hat{\rho}^L_{[0,T]} := \frac{1}{T}\int_0^T \hat{\rho}^L_t \,\mathrm{d}t$ and current $\hat{j}^L_{[0,T]} := \frac{1}{T}\int_0^T \hat{j}^L_t \,\mathrm{d}t$ in MFT (recall Sect. 4.3 for definitions) also satisfy a joint LDP in the limit $L, T \to \infty$: one takes first $L \to \infty$ and then $T \to \infty$, see [26, Eq. (36)]. The LDP is similar to (63):

$$\mathrm{Prob}\left(\left(\hat{\rho}^L_{[0,T]}, \hat{j}^L_{[0,T]}\right) \approx (\rho, j)\right) \asymp \exp\left\{-T|\Lambda_L| I^{\mathrm{MFT}}_{\mathrm{joint}}(\rho, j)\right\}, \tag{83}$$

where the rate function is, for a density profile ρ and a current j with $\mathrm{div}\, j = 0$, given by

$$I^{\mathrm{MFT}}_{\mathrm{joint}}(\rho, j) = \frac{1}{2}\Phi_{\mathrm{MFT}}(\rho, j, F(\rho)). \tag{84}$$

As for Markov chains (see Sect. 5.1) $I^{\mathrm{MFT}}_{\mathrm{joint}}(\rho, j) = \infty$ if j is not divergence free. If $\mathrm{div}\, j = 0$ then the rate function can be written in the form [26]

$$\begin{aligned} I^{\mathrm{MFT}}_{\mathrm{joint}}(\rho, j) &= \frac{1}{4}\int_\Lambda \nabla\frac{\delta\mathcal{V}}{\delta\rho}\cdot\chi\nabla\frac{\delta\mathcal{V}}{\delta\rho}\,\mathrm{d}x + \frac{1}{4}\int_\Lambda \nabla\varphi\cdot\chi\nabla\varphi\,\mathrm{d}x \\ &\quad + \frac{1}{4}\int_\Lambda (J_F - j)\cdot\chi^{-1}(J_F - j)\,\mathrm{d}x, \end{aligned} \tag{85}$$

such that a contraction to to the density only yields

$$I^{\mathrm{MFT}}_{\mathrm{density}}(\rho) = \frac{1}{4}\int_\Lambda \nabla\frac{\delta\mathcal{V}}{\delta\rho}\cdot\chi\nabla\frac{\delta\mathcal{V}}{\delta\rho}\,\mathrm{d}x + \frac{1}{4}\int_\Lambda \nabla\varphi\cdot\chi\nabla\varphi\,\mathrm{d}x. \tag{86}$$

The function φ in (85) and (86) is obtained by solving

$$\mathrm{div}\, J_F(\rho) = 0, \qquad J_F(\rho) := \chi\nabla\varphi + J_A(\rho). \tag{87}$$

Clearly the solution φ depends on ρ. In essence, we have reduced the minimisation problem (80) to the solution of this PDE. Comparing with (78), we identify the terms $J_F = \chi\tilde{F}^A$ in the MFT setting, and also $\tilde{J}^{\mathrm{ss}} = \chi\tilde{F}^A$, so $(\tilde{J}^{\mathrm{ss}} - \chi F^A(\rho)) = \chi\nabla\varphi$. We obtain the following representations for (85) and (86) reminiscent of Proposition 8 and Corollary 9.

Proposition 10 *The rate functional for the joint density and current in MFT, which is given by* (85)*, can be written in terms of the OM functional* (51) *as*

$$I^{\mathrm{MFT}}_{\mathrm{joint}}(\rho, j) = \frac{1}{2}\Big[\Phi_{\mathrm{MFT}}(\rho, 0, F^S(\rho)) + \Phi_{\mathrm{MFT}}(\rho, \tilde{J}^{\mathrm{ss}}, F^A(\rho)) + \Phi_{\mathrm{MFT}}(\rho, j, \tilde{F}^A)\Big], \tag{88}$$

and (86)*, the rate functional for the density in MFT, is given by*

$$I^{\mathrm{MFT}}_{\mathrm{density}}(\rho) = \frac{1}{2}\Big[\Phi_{\mathrm{MFT}}(\rho, 0, F^S(\rho)) + \Phi_{\mathrm{MFT}}(\rho, \tilde{J}^{\mathrm{ss}}, F^A(\rho))\Big]. \tag{89}$$

This proposition is equivalent to Proposition 5 of [26], but has now been rewritten in the language of optimal control theory. As discussed in [26], Eq. (89) quantifies the extent to which breaking detailed balance accelerates convergence of systems to equilibrium, at the hydrodynamic level. For this work, the key point is that this result originates from Corollary 9, which is the equivalent statement for Markov chains (without taking any hydrodynamic limit).

6 Consequences of the Structure of the OM Functional Φ

We have shown that the rate functions for several LDPs in several different contexts depend on functionals Φ with the general structure presented in (3) and (4). In this section, we show how this structure alone is sufficient to establish some features that are well-known in MFT. This means that these results within MFT have analogues for Markov chains. Our derivations mostly follow the standard MFT routes [7], but we use a more abstract notation to emphasise the minimal assumptions that are required.

6.1 Assumptions

The following minimal assumptions are easily verified for Markov chains; they are also either assumed or easily proven for MFT. The results of this section are therefore valid in both settings.

We consider a process described by a time-dependent density ρ and current j, with an associated continuity equation $\dot{\rho} = -\operatorname{div} j$ and unique steady state π. We are given a set of (ρ-dependent) forces denoted by $F(\rho)$, a dual pairing $j \cdot f$ between forces and currents, and a function $\Psi(\rho, j)$ which is convex in j and satisfies $\Psi(\rho, j) = \Psi(\rho, -j)$. With these choices, the functions $\Psi^\star$ and Φ are fully specified via (3) and (4). We assume that for initial conditions chosen from the invariant measure, the system satisfies an LDP of the form (1) with rate function of the form (2).

We define an adjoint process for which the probability of a path $(\rho_t, j_t)_{t\in[0,T]}$ is equal to the probability of the time-reversed path $(\rho^*_t, j^*_t)_{t\in[0,T]}$ in the original process. As above, we define $(\rho^*_t, j^*_t) = (\rho_{T-t}, -j_{T-t})$. We assume that the adjoint process also satisfies an LDP of the form (1), with rate function $I^*_{[0,T]}$. Hence we must have

$$I^*_{[0,T]}\big((\rho_t, j_t)_{t\in[0,T]}\big) = I_{[0,T]}\big((\rho^*_t, j^*_t)_{t\in[0,T]}\big). \tag{90}$$

Moreover, we assume that $I^*_{[0,T]}$ may be obtained from I by replacing the force $F(\rho)$ with some adjoint force $F^*(\rho)$. That is,

$$I^*_{[0,T]}\big((\rho_t, j_t)_{t\in[0,T]}\big) = I_0(\rho_0) + \frac{1}{2}\int_0^T \Phi(\rho_t, j_t, F^*(\rho_t))\,\mathrm{d}t. \tag{91}$$

Here, I_0 is the rate function associated with fluctuations of the density ρ, for a system in its steady state. That is, within the steady state, $\mathrm{Prob}(\hat{\rho}^{\mathcal{N}} \approx \rho) \asymp \exp(-\mathcal{N} I_0(\rho))$. For Markov chains, $I_0 = \mathcal{F}$, the free energy; for MFT we have $I_0 = \mathcal{V}$, the quasipotential. In the following we refer to I_0 as the free energy.

6.2 Symmetric and Anti-symmetric Forces

Define

$$F^S(\rho) := \frac{1}{2}[F(\rho) + F^*(\rho)], \qquad F^A(\rho) := \frac{1}{2}[F(\rho) - F^*(\rho)]. \tag{92}$$

As the following proposition shows, F^S is connected to the gradient of the free energy (or quasipotential) I_0, and the forces F^A and F^S satisfy a generalised orthogonality (in the sense of Proposition 3). The proof follows Section II.C of [7], but uses only the assumptions of Sect. 6.1, showing that the result applies also to Markov chains.

Proposition 11 *The forces F^S and F^A satisfy*

$$F^S(\rho) = -\nabla \frac{\delta I_0}{\delta \rho}, \tag{93}$$

and

$$\Psi^\star\big(\rho, F^S(\rho) + F^A\big) = \Psi^\star\big(\rho, F^S(\rho) - F^A\big). \tag{94}$$

Proof Combining (90) and (91), we obtain (for any path $(\rho_t, j_t)_{t\in[0,T]}$ that obeys the continuity equation $\dot\rho = -\operatorname{div} j$)

$$I_0(\rho_0) + \frac{1}{2}\int_0^T \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t = I_0(\rho_T) + \frac{1}{2}\int_0^T \Phi(\rho_{T-t}, -j_{T-t}, F^*(\rho_{T-t}))\,\mathrm{d}t. \tag{95}$$

Differentiating with respect to T and using (3) together with $\Psi(\rho, j) = \Psi(\rho, -j)$ and (92), one has

$$\dot I_0(\rho) + j \cdot F^S(\rho) + \frac{1}{2}\big[\Psi^\star(\rho, F^*(\rho)) - \Psi^\star(\rho, F(\rho))\big] = 0.$$

Using the continuity equation and an integration by parts, one finds $\dot I_0(\rho) = j \cdot \nabla \frac{\delta I_0}{\delta \rho}$, so that

$$j \cdot \left[F^S(\rho) + \nabla \frac{\delta I_0}{\delta \rho}\right] + \frac{1}{2}\big[\Psi^\star(\rho, F^*(\rho)) - \Psi^\star(\rho, F(\rho))\big] = 0.$$

This equation must hold for all (ρ, j), which means that the two terms in square parentheses both vanish separately. Combining the last equation with (92), we obtain (93) and (94). □

Proposition 11 also yields a variational characterisation of I_0. The following corollary is analogous to Eq. (4.8) of [7], as is its proof.

Corollary 12 *The free energy I_0 satisfies*

$$I_0(\hat\rho) = \inf \frac{1}{2}\int_{-\infty}^0 \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t, \tag{96}$$

where the infimum is taken over all paths $(\rho_t, j_t)_{t\in(-\infty,0]}$ that satisfy $\dot\rho_t + \operatorname{div} j_t = 0$, as well as $\lim_{t\to-\infty}\rho_t = \pi$ and $\rho_0 = \hat\rho$. Moreover, the optimal path is given by the time reversal of the solution of the adjoint dynamics $(\rho_t, -J^*(\rho_t))_{t\in(-\infty,0]}$.

Proof We obtain from (95) (together with (2) and (90)) that

$$\frac{1}{2}\int_{-\infty}^0 \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t = I_0(\hat\rho) + \frac{1}{2}\int_{-\infty}^0 \Phi(\rho_t, j_t, F^*(\rho_t))\,\mathrm{d}t.$$

Taking the infimum on both sides yields (96); indeed the infimum of $\frac{1}{2}\int_{-\infty}^0 \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t$ is 0, and this infimum is attained uniquely for the optimal

path for (96). To see this, we note that $\Phi(\rho_t, -j_t, F^*(\rho_t))$ is uniquely minimised for $j_t = -J^*(\rho_t)$, and $(\rho_t, -J^*(\rho_t))_{t\in(-\infty,0]}$ satisfies the conditions above, so the optimal path is indeed the time-reversal of the solution of the adjoint dynamics. □

6.3 Hamilton–Jacobi Like Equation for the Extended Hamiltonian

Another important relationship within MFT is the Hamilton–Jacobi equation [7, Eq. (4.13)]. This provides a characterisation of the quasipotential, as its maximal non-negative solution. The following formulation of that result uses only the assumptions of Sect. 6.1 and therefore applies also to Markov chains. The functional

$$\mathbb{L}(\rho, j) := \frac{1}{2}\Phi(\rho, j, F(\rho)) \tag{97}$$

can be interpreted as an extended Lagrangian. (Note that $\mathbb{L}(\rho, j)$ should not be interpreted as a Lagrangian in the classical sense, as it depends on density and current (ρ, j), rather than the pair consisting of density and associated velocity $(\rho, \dot\rho)$). We follow Sect. IV.G of [7]: given a sample path $(\rho_t, j_t)_{t\in[0,T]}$), define a vector field $A_t = A_0 - \int_0^t j_s \mathrm{d}s$. The initial condition A_0 is chosen so that there is a bijection between the paths $(\rho_t, j_t)_{t\in[0,T]}$ and $(A_t)_{t\in[0,T]}$. For example, in finite Markov chains, define $\bar\rho$ as a constant density, normalised to unity, and let $A_0 = \nabla h$, where h solves $\mathrm{div}(\nabla h) = (\rho_0 - \bar\rho)$, see [13] for the relevant properties of these vector fields. With this choice, and using $\dot\rho = -\,\mathrm{div}\, j$, one has $\rho_t = \bar\rho + \mathrm{div}\, A_t$ for all t, and one may also write (formally) $A_t = \mathrm{div}^{-1}(\rho_t - \bar\rho)$. Comparing with [7, Sect. IV.G], we write $\rho = \bar\rho + \mathrm{div}\, A$ instead of $\rho = \mathrm{div}\, A$ since for Markov chains one has (for any discrete vector field A) that $\sum_x \mathrm{div}\, A(x) = 0$, so it is not possible to solve $\mathrm{div}\, A = \rho$ if ρ is normalised to unity (recall that discrete vector fields have by definition $A_{xy} = -A_{yx}$ [13]).

The fluctuations of A are therefore determined by the fluctuations of (ρ, j), so the LDP (1) implies a similar LDP for A, whose rate function is $I^{\mathrm{ex}}_{[0,T]}((A_t)_{t\in[0,T]}) = I^{\mathrm{ex}}_0(A_0) + \int_0^T \mathbb{L}^{\mathrm{ex}}(A_t, \dot A_t)\mathrm{d}t$, where $\mathbb{L}^{\mathrm{ex}}$ is a Lagrangian that depends on A and its time derivative (which we again refer to as extended Lagrangian, cf. [7]). The function $\mathbb{L}$ in (97) is then related to $\mathbb{L}^{\mathrm{ex}}$ via the bijection between (ρ, j) and A. Considering again the case of Markov chains, the time evolution of the system depends only on $\mathrm{div}\, A$ (which is $\rho - \bar\rho$) and not on A itself, one sees that $\mathbb{L}^{\mathrm{ex}}(A, \dot A)$ depends only on $\mathrm{div}\, A$ and $\dot A$ (which is j). Hence we write, formally, $\mathbb{L}(\rho, j) = \mathbb{L}^{\mathrm{ex}}(\mathrm{div}^{-1}(\rho - \bar\rho), -j)$, and we recover (97).

Hence $\mathbb{L}$ is nothing but the extended Lagrangian $\mathbb{L}^{\mathrm{ex}}$, written in different variables: for this reason we refer to $\mathbb{L}$ as an (extended) Lagrangian.

To arrive at the corresponding (extended) Hamiltonian, one should write $\mathbb{H}^{\mathrm{ex}}(A, \xi) = \sup_{\dot A}[\xi \cdot \dot A - \mathbb{L}^{\mathrm{ex}}(A_t, \dot A_t)]$, or equivalently

$$\mathbb{H}(\rho, \xi) = \sup_j \big(j \cdot \xi - \mathbb{L}(\rho, j)\big), \tag{98}$$

where ξ is a conjugate field for the current j. We identify $\mathbb{H}$ as the scaled cumulant generating function associated with the rate function $I_{2.5}(\rho, j) = \mathbb{L}(\rho, j)$ [52, Sect. 3.1]. Analysis of rare fluctuations in terms of the field ξ is often more convenient than direct analysis of the rate function [32,33] and is the basis of the "s-ensemble" method that has recently been exploited in a number of physical applications (for example [21,24]). Using (3) and (4), we obtain

$$\mathbb{H}(\rho, \xi) = \frac{1}{2}\Psi^\star(\rho, F(\rho) + 2\xi) - \frac{1}{2}\Psi^\star(\rho, F(\rho)). \tag{99}$$

(This generalises the definition (44), which was restricted to Markov chains.)

To relate this extended Hamiltonian to the free energy (quasipotential), one can define an *extended Hamilton–Jacobi equation*, which is for a functional $\mathcal{S}$ given by

$$\mathbb{H}\left(\rho, \nabla\frac{\delta\mathcal{S}}{\delta\rho}\right) = 0. \tag{100}$$

The relation of this equation to the free energy is given by the following proposition, which mirrors equation (4.18) of [7], but now in our generalised setting, so that it applies also to Markov chains.

Proposition 13 *The free energy I_0 is the maximal non-negative solution to (100) which vanishes at the steady state π. In other words, any functional $\mathcal{S}$ that solves (100) and has $\mathcal{S}(\pi) = 0$ also satisfies $\mathcal{S} \le I_0$.*

Proof From (92), (93), (94) and $\Psi^\star(\rho, F) = \Psi^\star(\rho, -F)$, one has

$$\Psi^\star\left(\rho, F(\rho) + 2\nabla\tfrac{\delta I_0}{\delta\rho}\right) = \Psi^\star(\rho, -F_S(\rho) + F_A(\rho)) = \Psi^\star(\rho, F(\rho)). \tag{101}$$

Thus (99) yields $\mathbb{H}\big(\rho, \nabla\frac{\delta I_0}{\delta\rho}\big) = 0$, so I_0 does indeed solve (100). In addition, (101) is valid also with I_0 replaced by any $\mathcal{S}$ that solves (100); combining this result with (3) yields

$$\Phi(\rho, j, F(\rho)) = \Phi\left(\rho, j, F(\rho) + 2\nabla\frac{\delta\mathcal{S}}{\delta\rho}\right) + 2j\cdot\nabla\frac{\delta\mathcal{S}}{\delta\rho} \ge 2j\cdot\nabla\frac{\delta\mathcal{S}}{\delta\rho}, \tag{102}$$

where the second step uses $\Phi \ge 0$. Moreover, for any path $(\rho_t, j_t)_{t\in(-\infty,0]}$ with $\dot\rho_t + \mathrm{div}\, j_t = 0$ and $\lim_{t\to-\infty}\rho_t = \pi$, we have from (102) that

$$\begin{aligned} I_{(-\infty,0]}\big((\rho, j)_{t\in(-\infty,0]}\big) &= \int_{-\infty}^{0} \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t \\ &\ge \int_{-\infty}^{0} j(x)\cdot\nabla\frac{\delta\mathcal{S}}{\delta\rho}(x)\,\mathrm{d}t = \mathcal{S}(\rho_0), \end{aligned}$$

where the final equality uses an integration by parts, together with the continuity equation. Finally, taking the infimum over all paths and using Corollary 12, one obtains $\mathcal{S}(\rho) \le I_0(\rho)$, as claimed. □

6.4 Generalisation of Lemma 2

Before ending, we note that (94) is analogous to Proposition 3 in the general setting of this section, but we have not yet proved any analogue of Lemma 2. Hence we have not obtained a generalisation of Corollary 4, nor any of its further consequences. To achieve this, one requires a further assumption within the general framework considered here, which amounts to a splitting of the Hamiltonian. This assumption holds for MFT and for Markov chains, and is a sufficient condition for a generalised Lemma 2.

To state the assumption, we consider a reversible process in which the forces are $F^S(\rho)$. (For Markov chains we should consider the process with rates $r^S_{xy} = \frac{1}{2}(r_{xy} + r^*_{xy})$; for MFT it is the process with $J(\rho) = J^S(\rho)$ and the same mobility χ as the original process.) We assume that such a process exists and that its Hamiltonian can be written as $\mathbb{H}_S(\rho,\xi) = \frac{1}{2}[\Psi^\star_S(\rho, F^S(\rho)+2\xi) - \Psi^\star_S(\rho, F^S(\rho))]$ for some function $\Psi^\star_S$ (compare (99) and see Sect. 3.4 for the case of Markov chains). Also let the Hamiltonian for the adjoint process be $\mathbb{H}^*(\rho,\xi)$, which is constructed by replacing F by F^* in (99). Then, one assumes further that

$$\mathbb{H}_S(\rho,\xi) = \tfrac{1}{2}[\mathbb{H}(\rho,\xi) + \mathbb{H}^*(\rho,\xi)], \tag{103}$$

which may be verified to hold for Markov chains and for MFT. Writing $\xi = -F^S/2$ and using (99) with (94) and $\Psi^\star(\rho, f) = \Psi^\star(\rho, -f)$, one then obtains

$$\Psi^\star_S(\rho, F^S(\rho)) = \Psi^\star(F(\rho)) - \Psi^\star(F^A(\rho)), \tag{104}$$

which is the promised generalisation of Lemma 2.

7 Conclusion

In this article, we have presented several results for dynamical fluctuations in Markov chains. The central object in our discussion has been the function Φ, which plays a number of different roles—it is the rate function for large deviations at level 2.5 (Eq. 64), and it also appears in the rate function for pathwise large deviation functions (Eq. 2). These results—derived originally by Maes et al. [38,39]—originate from the relationship between Φ and the relative entropy between path measures (Appendix A). The canonical (Legendre transform) structure of Φ (Eq. 4) and its relation to time reversal (Eq. 25) have also been discussed before [38].

The function Φ depends on probability currents j and their conjugate forces f. Our Proposition 3 and Corollary 4 show how the rate functions in which Φ appears have another level of structure, based on the decomposition of the forces F in two pieces $F = F^S + F^A$, according to its behaviour under time-reversal. A similar decomposition is applied in MFT [7]: the discussion of Sects. 5 and 6 show how several results of that theory—which applies on macroscopic (hydrodynamic) scales—already have analogues for Markov chains, which provide microscopic descriptions of interacting particle systems. These results—which concern symmetries, gradient structures and (generalised) orthogonality relationships—show how properties of the rate functions are directly connected to physical ideas of free energy, dissipation, and time-reversal.

Looking forward, we hope that these structures can be exploited both in mathematics and physics. From a mathematical viewpoint, the canonical structure and generalised orthogonality relationships may provide new routes for scale-bridging calculations, just as the geometrical structure identified by Maas [35] has been used to develop new proofs of hydrodynamic limits [17]. In physics, a common technique is to propose macroscopic descriptions of physical systems based on symmetries and general principles—examples in non-equilibrium (active) systems include [51,54]. However, this level of description leaves some ambiguity as to the best definitions of some physical quantities, such as the local entropy production [44]. We hope that the structures identified here can be useful in relating such macroscopic theories to underlying microscopic behaviour.

Acknowledgements We thank Freddy Bouchet, Davide Gabrielli, Juan Garrahan, Jan Maas, Michiel Renger and Hugo Touchette for useful discussions. MK is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/L015684/1. JZ gratefully acknowledges funding by the EPSRC through project EP/K027743/1, the Leverhulme Trust (RPG-2013-261) and a Royal Society Wolfson Research Merit Award. The authors thank the anonymous referees for their careful reading of the manuscript and for many helpful comments and suggestions.

Appendix A: Relative Entropy on Path Space

Consider a Markov process with rates $r(x, y)$ and initial distribution Q_0. We fix a time interval $[0, T]$ for some $T > 0$ and denote the distribution of the Markov process on this time interval with Q. For each path $(x_u)_{u\in[0,T]}$ with jumps at times $t_1, \dots, t_n$ the density of Q can be found by solving the associated master equation (11); it is given by

$$Q\big((x_u)_{u\in[0,T]}\big) = Q_0(x_0)\exp\left\{\int_0^T \left(\sum_{i=1}^n \log r_t(x_{t-}, x_t)\delta(t-t_i) - \sum_y r_t(x_t, y)\right)\mathrm{d}t\right\},$$

where $x_{t-} := \lim_{\epsilon\to 0} x_{t-\epsilon}$ is the state of the process just before time t.

Now consider a second Markov process with time-dependent rates $\hat{r}_t(x, y)$ and initial distribution P_0. The distribution of this process is denoted by P. The logarithmic density of P with respect to Q is given by

$$\begin{aligned}\log\frac{dP}{dQ}\big((x_u)_{u\in[0,T]}\big) &= \log\frac{dP_0}{dQ_0}(x_0)\\ &+ \int_0^T \left(\sum_{i=1}^n \log\Big(\frac{\hat{r}_t(x_{t^-}, x_t)}{r(x_{t^-}, x_t)}\Big)\delta(t-t_i) - \sum_y\big[\hat{r}_t(x_t, y) - r(x_t, y)\big]\right)\mathrm{d}t.\end{aligned}$$

We further denote the distribution of P at time t with ρ_t, such that $\rho_t = P\circ X_t^{-1}$ where X_t denotes the evaluation of the path at time t (such that in particular $P_0 = \rho_0$). The *relative entropy* on path space

$$\mathcal{H}(P|Q) := \mathbb{E}_P\left[\log\Big(\frac{dP}{dQ}\Big)\right]$$

is then equal to

$$\mathbb{E}_{P_0}\left[\log\Big(\frac{dP_0}{dQ_0}\Big)\right] + \int_0^T \sum_{x,y}\rho_t(x)\Big(\hat{r}_t(x, y)\log\Big(\frac{\hat{r}_t(x, y)}{r(x, y)}\Big) - \hat{r}_t(x, y) + r(x, y)\Big)\mathrm{d}t.$$

Let $(\rho_t, j_t)_{t\in[0,T]}$ be given, with $\rho_t > 0$ for all times $t \in [0, T]$. We then can rewrite the relative entropy $\mathcal{H}(P|Q)$ in terms of the flow $C_t(x, y) := \rho_t(x)\hat{r}_t(x, y)$ as

$$\mathcal{H}(\rho_0|Q_0) + \int_0^T \sum_{x,y}\Big(C_t(x, y)\log\Big(\frac{C_t(x, y)}{\rho_t(x)r(x, y)}\Big) - C_t(x, y) + \rho_t(x)r(x, y)\Big)\mathrm{d}t. \quad (105)$$

Note that the relative entropy $\mathcal{H}(P|Q)$ can (just as the Markov chain) be completely characterised by the probability distribution $(\rho_t)_{t\in[0,T]}$ and the flow $(C_t)_{t\in[0,T]}$.

We are interested in a special flow $(C_t)_{t\in[0,T]}$ which recovers a given current $(j_t)_{t\in[0,T]}$ as $(j_t)_{xy} = C_t(x, y) - C_t(y, x)$. The force associated to j_t is by (13) given by $f^{j_t}(\rho_t) := 2\,\mathrm{arcsinh}(j_t/a(\rho_t))$ and the flow of interest is defined as $C_t(x, y) = \frac{1}{2}a_{xy}(\rho_t)\exp(\frac{1}{2}f^{j_t}_{xy}(\rho_t))$. It can be interpreted as the optimal flow that creates the current $(j_t)_{t\in[0,T]}$.

We define the rates $\tilde{r}_t(x, y) := C_t(x, y)/\rho_t(x)$ and denote the law of the associated (time heterogeneous) Markov process on $[0, T]$ with $\tilde{P}$. The relative entropy of this new process $\tilde{P}$ with respect to the reference process Q is

$$\mathcal{H}(\tilde{P}|Q) = \mathcal{H}(\rho_0|Q_0) + \frac{1}{2}\int_0^T \Phi(\rho_t, j_t, F(\rho_t))\,\mathrm{d}t \quad (106)$$

with Φ given by (3); to see this, we argue as follows. Symmetrising (105) and considering each summand separately gives

$$\frac{1}{2}\Big(C_t(x,y)\log\frac{C_t(x,y)}{C_t^Q(x,y)}+C_t(y,x)\log\frac{C_t(y,x)}{C_t^Q(y,x)}\Big)$$
$$+\frac{1}{2}\Big(C_t^Q(x,y)-C_t(x,y)+C_t^Q(y,x)-C_t(y,x)\Big),$$

where the first summand coincides with

$$\frac{1}{2}\left(\frac{1}{2}a_{xy}(\rho_t)\sinh\big(\tfrac{1}{2}f_{xy}^{j_t}(\rho_t)\big)f_{xy}^{j_t}(\rho_t)-\frac{1}{2}a_{xy}(\rho_t)\sinh\big(\tfrac{1}{2}f_{xy}^{j_t}(\rho_t)\big)F_{xy}(\rho_t)\right)$$

and the second is given by

$$\frac{1}{2}\Big(a_{xy}(\rho_t)\cosh\big(\tfrac{1}{2}F_{xy}(\rho_t)\big)-a_{xy}(\rho_t)\cosh\big(\tfrac{1}{2}f_{xy}^{j_t}(\rho_t)\big)\Big).$$

Combining this with (15) and (16) yields (106).

Pathwise Large Deviation Principle: Let $x^1, x^2, \ldots$ be a sequence of iid copies of the Markov chains with law Q. By Sanov's Theorem (see, e.g., Theorem 6.2.10 in [14]), the empirical average $\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}\delta_{x^i}$ of the Markov chains satisfies a LDP with the rate functional $\mathcal{H}(\cdot|Q)$. We can interpret $\mathcal{H}(\cdot|Q)$ as the rate functional for the joint LDP of $(\rho_t, C_t)_{t\in[0,T]}$ by defining this rate functional $\mathcal{I}_{[0,T]}((\rho_t, C_t)_{t\in[0,T]})$ as the right-hand side of (105).

We contract the above rate functional to obtain the rate functional for the joint empirical measure and current $(\rho_t, j_t)_{t\in[0,T]}$. It is given by

$$I_{[0,T]}((\rho_t, j_t)_{t\in[0,T]}) := \inf_{(C_t)_{t\in[0,T]}} \mathcal{I}_{[0,T]}((\rho_t, C_t)_{t\in[0,T]}), \tag{107}$$

where the infimum is taken over the set of all flows which yield the current $(j_t)_{t\in[0,T]}$, i.e. over the set $\{(C_t)_{t\in[0,T]}|$ for all $t \in [0,T] : C_t(x,y) \geq 0$ and $C_t(x,y) - C_t(y,x) = (j_t)_{xy}\}$. It was shown in [38] and [9] that the minimising flow is the current $C_t(x,y) = \frac{1}{2}a_{xy}(\rho_t)\exp(\frac{1}{2}f_{xy}^{j_t}(\rho_t))$ introduced above, such that $I_{[0,T]}((\rho_t, j_t)_{t\in[0,T]})$ coincides with (106).

References

1. Andrieux, D., Gaspard, P.: Fluctuation theorem for currents and Schnakenberg network theory. J. Stat. Phys. **127**(1), 107–131 (2007)
2. Basile, G., Benedetto, D., Bertini, L.: A gradient flow approach to linear Boltzmann equations. arXiv preprint. arXiv:1707.09204 (2017)
3. Basu, U., Maes, C.: Nonequilibrium response and frenesy. J. Phys: Conf. Ser. **638**(1), 012001 (2015)
4. Benois, O., Kipnis, C., Landim, C.: Large deviations from the hydrodynamical limit of mean zero asymmetric zero range processes. Stoch. Process. Appl. **55**(1), 65–89 (1995)
5. Bernard, E.P., Krauth, W.: Two-step melting in two dimensions: first-order liquid-hexatic transition. Phys. Rev. Lett. **107**, 155704 (2011)
6. Bertini, L., Landim, C., Mourragui, M., et al.: Dynamical large deviations for the boundary driven weakly asymmetric exclusion process. Ann. Probab. **37**(6), 2357–2403 (2009)
7. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Macroscopic fluctuation theory. Rev. Mod. Phys. **87**(2), 593–636 (2015)
8. Bertini, L., Faggionato, A., Gabrielli, D.: Flows, currents, and cycles for Markov chains: large deviation asymptotics. Stoch. Process. Appl. **125**(7), 2786–2819 (2015)
9. Bertini, L., Faggionato, A., Gabrielli, D.: Large deviations of the empirical flow for continuous time Markov chains. Ann. Inst. Henri Poincaré Probab. Stat. **51**(3), 867–900 (2015)

10. Chernyak, V.Y., Chertkov, M., Bierkens, J., Kappen, H.J.: Stochastic optimal control as non-equilibrium statistical mechanics: calculus of variations over density and current. J. Phys. A **47**(2), 022001 (2014)
11. Chetrite, R., Touchette, H.: Variational and optimal control representations of conditioned and driven processes. J. Stat. Mech. Theory Exp. **2015**(12), P12001, 42 (2015)
12. Crooks, G.E.: Path-ensemble averages in systems driven far from equilibrium. Phys. Rev. E **61**, 2361–2366 (2000)
13. De Carlo, L., Gabrielli, D.: Gibbsian stationary non-equilibrium states. J. Stat. Phys. **168**(6), 1191–1222 (2017)
14. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications. Stochastic Modelling and Applied Probability, vol. 38. Springer, Berlin (2010). Corrected Reprint of 2nd edn (1998)
15. den Hollander, F.: Large Deviations. Fields Institute Monographs, vol. 14. American Mathematical Society, Providence (2000)
16. Esposito, M., Van den Broeck, C.: Three faces of the second law. I. Master equation formulation. Phys. Rev. E **82**, 011143 (2010)
17. Fathi, M., Simon, M.: The Gradient Flow Approach to Hydrodynamic Limits for the Simple Exclusion Process, pp. 167–184. Springer, Cham (2016)
18. Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions. Stochastic Modelling and Applied Probability, vol. 25, 2nd edn. Springer, New York (2006)
19. Gallavotti, G., Cohen, E.G.D.: Dynamical ensembles in stationary states. J. Stat. Phys. **80**(5–6), 931–970 (1995)
20. Gardiner, C.: Stochastic Methods. A Handbook for the Natural and Social Sciences. Springer Series in Synergetics, 4th edn. Springer, Berlin (2009)
21. Garrahan, J.P., Jack, R.L., Lecomte, V., Pitard, E., van Duijvendijk, K., van Wijland, F.: First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories. J. Phys. A **42**(7), 075007, 34 (2009)
22. Gingrich, T.R., Horowitz, J.M., Perunov, N., England, J.L.: Dissipation bounds all steady-state current fluctuations. Phys. Rev. Lett. **116**, 120601 (2016)
23. Gingrich, T.R., Rotskoff, G.M., Horowitz, J.M.: Inferring dissipation from current fluctuations. J. Phys. A **50**(18), 184004 (2017)
24. Jack, R.L., Sollich, P.: Effective interactions and large deviations in stochastic processes. Eur. Phys. J. Spec. Top. **224**(12), 2351–2367 (2015)
25. Jarzynski, C.: Nonequilibrium equality for free energy differences. Phys. Rev. Lett. **78**, 2690–2693 (1997)
26. Kaiser, M., Jack, R.L., Zimmer, J.: Acceleration of convergence to equilibrium in Markov chains by breaking detailed balance. J. Stat. Phys. **168**, 259–287 (2017)
27. Kipnis, C., Olla, S., Varadhan, S.R.S.: Hydrodynamics and large deviation for simple exclusion processes. Commun. Pure Appl. Math. **42**(2), 115–137 (1989)
28. Kipnis, C., Landim, C.: Scaling Limits of Interacting Particle Systems. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 320. Springer, Berlin (1999)
29. Kolomeisky, A.B., Fisher, M.E.: Molecular motors: a theorist's perspective. Annu. Rev. Phys. Chem. **58**(1), 675–695 (2007)
30. Kwon, C., Ao, P., Thouless, D.J.: Structure of stochastic dynamics near fixed points. Proc. Natl Acad. Sci. U.S.A. **102**(37), 13029–13033 (2005)
31. Lavorel, J.: Matrix analysis of the oxygen evolving system of photosynthesis. J. Theor. Biol. **57**(1), 171–185 (1976)
32. Lebowitz, J.L.: A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics. J. Stat. Phys. **95**(1–2), 333–365 (1999)
33. Lecomte, V., Appert-Rolland, C., van Wijland, F.: Thermodynamic formalism for systems with Markov dynamics. J. Stat. Phys. **127**(1), 51–106 (2007)
34. Ma, Y.-A., Fox, E.B., Chen, T., Wu, L.: A unifying framework for devising efficient and irreversible MCMC samplers. arXiv preprint. arXiv:1608.05973 (2016)
35. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. **261**(8), 2250–2292 (2011)
36. Machlup, S., Onsager, L.: Fluctuations and irreversible process. II. Systems with kinetic energy. Phys. Rev. **91**, 1512–1515 (1953)
37. Maes, C.: The fluctuation theorem as a Gibbs property. J. Stat. Phys. **95**(1–2), 367–392 (1999)
38. Maes, C., Netočný, K.: Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states. Europhys. Lett. EPL **82**(3), Art. 30003, 6 (2008)
39. Maes, C., Netočný, K., Wynants, B.: On and beyond entropy production: the case of Markov jump processes. Markov Process. Relat. Fields **14**(3), 445–464 (2008)
40. Maes, C., Netočný, K., Wynants, B.: Monotonicity of the dynamical activity. J. Phys. A **45**(45), 455001, 13 (2012)

41. Maes, C.: Netočný, Karel: Revisiting the Glansdorff-Prigogine criterion for stability within irreversible thermodynamics. J. Stat. Phys. **159**(6), 1286–1299 (2015)
42. Mariani, M.: A gamma-convergence approach to large deviations. arXiv preprint. arXiv:1204.0640 (2012)
43. Mielke, A., Peletier, M.A., Renger, D.R.M.: On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. Potential Anal. **41**(4), 1293–1327 (2014)
44. Nardini, C., Fodor, É., Tjhung, E., van Wijland, F., Tailleur, J., Cates, M.E.: Entropy production in field theories without time-reversal symmetry: quantifying the non-equilibrium character of active matter. Phys. Rev. X **7**, 021007 (2017)
45. Pietzonka, P., Barato, A.C., Seifert, U.: Universal bounds on current fluctuations. Phys. Rev. E **93**, 052145 (2016)
46. Polettini, M., Lazarescu, A., Esposito, M.: Tightening the uncertainty principle for stochastic currents. Phys. Rev. E **94**, 052104 (2016)
47. Qian, H.: A decomposition of irreversible diffusion processes without detailed balance. J. Math. Phys. **54**(5), 053302 (2013)
48. Renger, D.R.M.: Large deviations of specific empirical fluxes of independent Markov chains, with implications for macroscopic fluctuation theory. Weierstrass Institute. Preprint 2375 (2017)
49. Schnakenberg, J.: Network theory of microscopic and macroscopic behavior of master equation systems. Rev. Mod. Phys. **48**(4), 571–585 (1976)
50. Seifert, U.: Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys. **75**(12), 126001 (2012)
51. Toner, J., Yuhai, T.: Long-range order in a two-dimensional dynamical XY model: how birds fly together. Phys. Rev. Lett. **75**, 4326–4329 (1995)
52. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. **478**(1–3), 1–69 (2009)
53. Vaikuntanathan, S., Gingrich, T.R., Geissler, P.L.: Dynamic phase transitions in simple driven kinetic networks. Phys. Rev. E **89**, 062108 (2014)
54. Wittkowski, R., Tiribocchi, A., Stenhammar, J., Allen, R.J., Marenduzzo, D., Cates, M.E.: Scalar ϕ^4 field theory for active-particle phase separation. Nat. Commun. **5**, 4351 (2014)

3.2. Conclusions

We discussed a general structure for dynamical fluctuations of ergodic finite state Markov chains with a unique steady state π. In Section 2 we derived a Ψ-$\Psi^\star$ structure which is based on the non-linear flux-force relation $J(\rho) = a(\rho)\sinh(F(\rho)/2)$ and the thermodynamic force

$$F_{xy}(\rho) = \log \frac{\rho(x) r_{xy}}{\rho(y) r_{yx}}.$$

Given the steady state π, we decomposed (in Section 3) the force in a time reversal symmetric part $F^S_{xy}(\rho) = -\nabla^{x,y} \log(\rho/\pi)$, which gives rise to a gradient flow for the free energy $\mathcal{F}$, and an anti-symmetric part $F^A = F(\pi)$. Moreover, the time-reversed process coincides with the original process, where the force $F(\rho)$ is replaced with $F^*(\rho) = F^S(\rho) - F^A$, such that in particular $F^A = 0$ if and only if the detailed balance relation holds. This splitting of the force in $F^S(\rho)$ and F^A has a physical interpretation in terms of stochastic thermodynamics and the correctness of this splitting is reinforced by the 'generalised orthogonality' (cf. Proposition 3)

$$\Psi^\star(\rho, F^S(\rho) + F^A) = \Psi^\star(\rho, F^S(\rho) - F^A)$$

for the Ψ-$\Psi^\star$ structure associated to the non-linear flux-force relation, which was also related to an *extended Hamilton-Jacobi equation* (Section 3.4).

In Section 4 we discussed the Ψ-$\Psi^\star$ structure for MFT. In particular, we showed in Section 4.4 that the 'generalised orthogonality' is in this case equivalent to the classical orthogonality condition (30) in Chapter 2.

We also discussed in Section 4.5 the example of a weakly asymmetric simple exclusion process (WASEP) and outlined how the Ψ-$\Psi^\star$ structure in MFT arises in the hydrodynamic limit from the Ψ-$\Psi^\star$ structure for Markov chains. A rigorous derivation, at least in the case of time-reversal symmetric particle systems, will be the main topic of the next Chapter.

Chapter 4

A Variational Structure for Interacting Particle Systems and their Hydrodynamic Scaling Limits

In this chapter, we review the connection between interacting particles systems and their hydrodynamic scaling limits. We apply the Ψ-$\Psi^\star$ structure derived in Chapter 3 to time-reversal symmetric particle systems with *gradient dynamics*, such as the SEP and the ZRP. In particular, we show how the quadratic structure of MFT can be recovered from the non-quadratic Ψ-$\Psi^\star$ structure in the limit as the system size tends to infinity. The following preprint, which is again joint work with Robert L. Jack and Johannes Zimmer, is available on the arXiv [30].

4.1. Outline of the Article

In the following article we apply the variational (or canonical) structure from Chapter 3 to hydrodynamic scaling limits of a class of time-reversal symmetric particle systems (as we prior discussed in Section 2 in Chapter 2 and Section 4 in Chapter 3).

Given a particle system (e.g. the ZRP) we denote with P_L^V the distribution of the particle system on the space of càdlàg paths $\mathcal{D}$ on a fixed time interval $[0,T]$ (which we in the following refer to as *path measures*). Further, consider a second particle system (with the same initial condition and possibly different path measure P_L), for which we can characterise the deviation from P_L^V in terms of the *microscopic action* (the relative entropy $\mathcal{H}(P_L|P_L^{\tilde{V}})$). The latter can be represented in terms of the Ψ-$\Psi^\star$ structure discussed in Chapter 3 as

$$\mathcal{H}(P_L|P_L^V) = \frac{1}{2}\bigg[\mathcal{F}_{L,\alpha}^V(\mu_t^L) - \mathcal{F}_{L,\alpha}^V(\mu_0^L) + \int_0^T \big[\Psi_L(\mu_t^L, j_t^L) + \Psi_L^\star\big(\mu_t^L, F^V(\mu_t^L)\big)\big]\,\mathrm{d}t\bigg].$$

Now, under the assumption that P_L concentrates in the limit on a path $(\pi_t)_{t\in[0,T]}$, a natural question to ask is whether the non-quadratic Ψ-$\Psi^\star$ structure converges in the asymptotic limit to the quadratic structure known from MFT (cf. Section 4 in Chapter 3). More precisely, we investigate the question under which assumptions the microscopic action (and the individual contributions) converge to the macroscopic action

$$\begin{aligned}\mathbb{A}\big((\pi_t)_{t\in[0,T]}\big) = \frac{1}{2}\big[\mathcal{F}_\alpha^V(\rho_T) - \mathcal{F}_\alpha^V(\rho_0)\big] \\ + \frac{1}{4}\int_0^T \big[\|\dot\rho_t\|_{-1,\chi(\rho_t)}^2 + \|\Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\|_{-1,\chi(\rho_t)}^2\big]\,\mathrm{d}t.\end{aligned}$$

In Section 2 we define the class of particle systems and the microscopic and macroscopic quantities of interest. In particular, we discuss the *local equilibrium assumption*, which is a crucial part in the proof of scaling limits for interacting particle systems. The new convergence results are presented in Section 3. Sections 4 and 5 contain the proofs and additional technical material.

Appendix B: Statement of Authorship

<table>
<tr><th colspan="10">This declaration concerns the article entitled:</th></tr>
<tr><td colspan="10">A Variational Structure for Interacting Particle Systems and their Hydrodynamic Scaling Limits</td></tr>
<tr><td colspan="10">Publication status (tick one)</td></tr>
<tr><td>draft manuscript</td><td></td><td>Submitted</td><td>X</td><td>In review</td><td></td><td>Accepted</td><td></td><td>Published</td><td></td></tr>
<tr><td>Publication details (reference)</td><td colspan="9">Preprint: arXiv:1805.01411
Authors: Marcus Kaiser, Robert L. Jack and Johannes Zimmer</td></tr>
<tr><td>Candidate's contribution to the paper (detailed, and also given as a percentage).</td><td colspan="9">The bulk of the calculations have been performed by
the author of the thesis (70%).
All authors contributed equally to the presentation
of the content (33%).</td></tr>
<tr><td>Statement from Candidate</td><td colspan="9">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td colspan="5"></td><td>Date</td><td colspan="3">10.7.2018</td></tr>
</table>

A Variational Structure for Interacting Particle Systems and their Hydrodynamic Scaling Limits

Marcus Kaiser[1], Robert L. Jack[2,3,4], and Johannes Zimmer[1]

[1]Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK
[2]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge, Cambridge CB3 0WA, UK
[3]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[4]Department of Physics, University of Bath, Bath BA2 7AY, UK

Abstract

We consider hydrodynamic scaling limits for a class of reversible interacting particle systems, which includes the symmetric simple exclusion process and certain zero-range processes. We study a (non-quadratic) microscopic action functional for these systems. We analyse the behaviour of this functional in the hydrodynamic limit and we establish conditions under which it converges to the (quadratic) action functional of Macroscopic Fluctuation Theory. We discuss the implications of these results for rigorous analysis of hydrodynamic limits.

1 Introduction

Recently, a *canonical structure* has been introduced [29, 30] to describe dynamical fluctuations in stochastic systems. The resulting theory has several attractive features: Firstly, it applies to a wide range of systems, including finite-state Markov chains and Macroscopic Fluctuation Theory (MFT) [5], see [21]. Secondly, it is based on an *action functional* which is a relative entropy between probability measures on path spaces — this means that it provides a variational description of the systems under consideration, and the action can be related to large deviation rate functionals. Thirdly, it extends the classical Onsager-Machlup theory [34] in a natural way, by replacing the quadratic functionals that appear in that theory with a pair of convex but *non-quadratic* Legendre duals Ψ and $\Psi^\star$. (This is sometimes called a Ψ-$\Psi^\star$ representation [31].) In Onsager-Machlup theory and in MFT, the minimiser of the action describes the most probable evolution of a macroscopic system, either in terms of thermodynamic forces and fluxes (in Onsager-Machlup theory) or densities and fluxes (in MFT): this feature is maintained in the canonical structure.

This structure can be applied to any finite-state Markov chain and provides a unifying formulation of a wide range of systems [21]. In particular, lattice systems of interacting particles can be described by canonical structures in two ways: either on the microscopic (Markov chain) level via non-quadratic Legendre duals, or as a coarse-grained version through the hydrodynamic limit, where the action reduces to a quadratic MFT functional. One therefore expects that in the hydrodynamic scaling limit, the microscopic (non-quadratic) structure should converge (in some suitable sense) to the macroscopic one. Such a convergence would offer a new way to understand and derive hydrodynamic limits. The main question of this article is whether this natural conjecture holds.

We give a partial (positive) answer, by proving several theorems that relate the microscopic and macroscopic action functionals for interacting particle systems. Specifically, we consider a class of

systems on periodic lattices with *gradient dynamics* and a conserved number of particles, which includes as special cases the symmetric simple exclusion process and a large class of reversible zero-range processes. In the hydrodynamic limit, the number of lattice sites and the number of particles go to infinity together, at fixed density, and the microscopic transition rates have a parabolic scaling. (These are among the simplest models for which one can rigorously establish a hydrodynamic limit [22].)

Our analysis is based on the microscopic action, which is a relative entropy between two probability measures: one measure encodes the dynamics of the particle system itself (the *reference process*) and the other represents some other *observed process*, which is to be compared with the reference process. We consider observed processes that concentrate (in the hydrodynamic limit) on deterministic paths: in this case we show that the individual contributions to the macroscopic action are asymptotically dominated by their microscopic counterparts, see Theorem 3.4. Then, for a specific choice of the observed process (which is related to the hydrodynamic limit of the reference process), we show that the microscopic action converges to the macroscopic one, see Theorems 3.5 and 3.6.

The inspiration for this study comes from [18] and [16], which derive hydrodynamic (or mean-field) limits as minimisers of macroscopic action functionals, for the simple exclusion process [18] and for a McKean-Vlasov equation on a finite graph [16]. In common with these works, our approach is (loosely) based on the Sandier-Serfaty approach [37] to study sequences of gradient flows via Γ-convergence. However, our approach is different from [18, 16] because it starts from the (non-quadratic) canonical structure, instead of the *quadratic* structure for time-reversal symmetric Markov chains, that was independently derived by Maas [28] and Mielke [32]. A similar structure to the canonical one exploited here was recently used in [2] to derive a diffusive limit for the linear Boltzmann equation. All of these approaches have in common that they consider time-reversal symmetric systems for which the dynamics can be identified with gradient flows of a free energy functional, so that the limiting probability measure concentrates on curves of maximal slope, which can be identified as minimisers of the macroscopic action. Further, our approach is also closely related to EDP-convergence, where EDP stands for Energy-Dissipation-Principle, see e.g. [24, 8, 14, 33].

Compared with previous studies, our work has two novel features. First, we do not restrict to curves of maximal slope (which follow the gradient of the free energy): instead we consider a class of paths for which the microscopic action functional stays controlled, in the hydrodynamic limit. In principle, this means that our methods are not limited to time-reversal symmetric systems: the corresponding action functional can be defined for a large class of Markov chains in a meaningful way. However, in order to reduce the number of technical issues we have to deal with, we limit ourselves to reversible systems in this work.

The second novel aspect is that we consider particle systems for which the hydrodynamic limit is a *non-linear* diffusion equation, in contrast (for example) to the symmetric exclusion process studied in [18], whose hydrodynamic limit is linear diffusion. This is a significant difference for rigorous results: within the canonical structure one sees naturally that the hydrodynamic limit is a (generalised) gradient flow, as expected on physical grounds. However, in contrast to (linear) diffusion with a linear mobility, where the (now-)classic Wasserstein evolution provides the natural geometrical setting for the gradient flow, the analogous setting for diffusions with non-linear mobility is not so well-developed. In particular, a key challenge is to establish the validity of a chain rule for the macroscopic entropy functional, which is known for linear diffusion [1], but whose extension to the non-linear setting is not at all straightforward. We show here that (with some technical effort) the required results for non-linear diffusion can be obtained by casting the evolution into the classic Wasserstein setting (Theorem 4.2): this is not the most natural (physical) setting for the process of interest, but it is sufficient to establish the required results.

This line of research — linking Markov chains and partial differential equations via canonical struc-

tures — is quite recent. Consequently, a number of problems remain open. In particular, our approach is not yet a hydrodynamic limit passage: for this, the macroscopic concentration of the limiting path measure would have to be proved. Also, the microscopic action converges in the hydrodynamic limit to a macroscopic action functional that turns out to coincide with a large deviation rate functional [5]. However, in this work we do not establish any links to large deviation theory; this could be a natural future line of research (e.g. one could consider similar calculations to the ones in [15] for independent particles with Langevin dynamics). Another question is whether (and how) the method presented here can provide guidance for limit passages for non-reversible systems.

Our study combines techniques from a number of different fields: we have attempted to make it self-contained (and hence accessible to a general reader), at the expense of including some classical material (which expert readers may prefer to skip). This is indicated in the beginning of the relevant sections. In Section 2, we describe the particle systems and their canonical structure. Section 3 states the main results. Section 4 is entirely devoted to technical questions of regularity and a proof of the chain rule, while Section 5 contains the proofs of the main theorems.

2 Interacting Particle Systems

2.1 Particle Systems on the Discrete Torus

The setting we analyse covers a broad class of particle models, as we now describe. This section also collects some classic facts on particle models. We consider systems with a fixed number of indistinguishable particles, distributed over the L^d sites of the flat torus $\mathbb{T}_L^d := \mathbb{Z}^d/(L\mathbb{Z}^d)$. Let $\eta(i)$ be the number of particles on site $i \in \mathbb{T}_L^d$, so the configuration space of the system is $\Omega_L \subseteq \mathbb{N}_0^{\mathbb{T}_L^d}$. Configurations are denoted with $\eta = (\eta(i))_{i\in\mathbb{T}_L^d}$. Let $\eta^{i,i'}$ be the configuration obtained from η by moving a particle from site i to site i'. The total number of particles on each site may be bounded by $N_{\max} \in \mathbb{N}_0$, that is, $\Omega_L = \{0,\dots,N_{\max}\}^{\mathbb{T}_L^d}$, or unbounded. We fix $T > 0$ and consider the time interval $[0,T]$. The (random) state of the system at time $t \in [0,T]$ is denoted by η_t.

The particles hop between sites of the lattice with some rate $\hat{r}_{\eta,\eta^{i,i'}}$, which is assumed to be non-zero only if i and i' are neighbours, $|i-i'| = 1$. We consider a parabolic scaling, so the hydrodynamic limit is obtained by rescaling time by a factor L^2, such that the transition rates for the Markov chain are $r_{\eta,\eta^{i,i'}} = L^2\hat{r}_{\eta,\eta^{i,i'}}$. Let Λ be the flat torus $\mathbb{T}^d = [0,1)^d$. The jump rates for the particle models considered in this article depend on an external potential $V \in C^2(\Lambda;\mathbb{R})$, and two functions $g_1, g_2 \colon \mathbb{N}_0 \to [0,\infty)$, such that

$$\hat{r}^V_{\eta,\eta^{i,i'}} = g_1(\eta(i))g_2(\eta(i'))\mathrm{e}^{-\frac{1}{2}(V(i'/L)-V(i/L))}. \tag{1}$$

We also consider time-dependent potentials $\tilde{V} \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ which lead to a time-heterogeneous Markov chain with transition rates $r^{\tilde{V}_t}$ at time $t \in [0,T]$. We write $\tilde{V}$ for a time-dependent potential and V for a time-independent potential.

An interacting particle system has *gradient dynamics* (or is of *gradient type*) if there exists a function $\mathrm{d}\colon \mathbb{N}_0 \to [0,\infty)$ such that (for $V = 0$) $r^0_{\eta,\eta^{i,i'}} - r^0_{\eta,\eta^{i',i}} = \mathrm{d}(\eta(i)) - \mathrm{d}(\eta(i'))$. In this case we define $\hat{\phi}_i(\mu) := \sum_{\eta\in\Omega_L}\mu(\eta)\,\mathrm{d}(\eta(i))$. (Note that this is the simplest form of a gradient system, which in more generality can consist of differences of finite cylinder functions, cf. [22]).

2.1.1 Invariant Measures, Initial Conditions, and Microscopic Free Energy

The number of particles is conserved by the dynamics, so these systems have many possible invariant measures. The hydrodynamic limit relies on a particular structure for these measures, as follows. Let

ν_* be a (not necessarily normalised) reference measure on Ω_L, with $\nu_*(\eta) > 0$ for all $\eta \in \Omega_L$, which is assumed to have a product structure in the sense that $\nu_*(\eta) = \prod_{i\in\mathbb{T}_L^d} \nu_{*,1}(\eta(i))$ for some probability measure $\nu_{*,1}$ on $\mathbb{N}_0$. We assume that the process with rates $\hat{r}^0$ satisfies the *detailed balance condition*

$$\nu_*(\eta)\,\hat{r}^0_{\eta,\eta^{i,i+e_k}} = \nu_*(\eta^{i,i+e_k})\,\hat{r}^0_{\eta^{i,i+e_k},\eta} \tag{2}$$

for all $\eta \in \Omega_L$, $i \in \mathbb{T}_L^d$ and $k = 1, \dots, d$. This implies that ν_* is invariant for the dynamics $\hat{r}^0$ and that these dynamics are time reversal-symmetric with respect to ν_*. To avoid technical difficulties, we further assume that the one site *partition function* is finite, i.e. for all $\theta \in \mathbb{R}$

$$Z_1(\theta) := \sum_{n\in\mathbb{N}_0} \mathrm{e}^{\theta n}\nu_{*,1}(n) < \infty. \tag{3}$$

In classical statistical mechanics (see for example [4, Section 3] or [9]), the *local free energy density* is given by the Legendre dual of the *cumulant generating function (or pressure)* of $\nu_{*,1}$, i.e.

$$f(a) = \sup_{\theta\in\mathbb{R}} \big(a\,\theta - \log Z_1(\theta)\big) = a\,f'(a) - \log Z_1(f'(a)), \tag{4}$$

which implies that f is convex. In the following, we will assume that $f \in C^2([0, N_{\max}]; \mathbb{R})$ and that a.e. $f'' > 0$, see Section 2.4.2). Now, for $\alpha \in (0, N_{\max})$, we define the probability measures

$$\nu_{\alpha,1}(n) := \frac{\mathrm{e}^{f'(\alpha)n}}{Z_1(f'(\alpha))}\nu_{*,1}(n) \tag{5}$$

and $\nu_\alpha := \prod_{i\in\mathbb{T}_L^d} \nu_{\alpha,1}$. For each $\alpha \in (0, N_{\max})$ this choice implies that $E_{\nu_\alpha}\big[\sum_{i\in\mathbb{T}_L^d} \eta(i)/L^d\big] = \alpha$ (where E_{ν_α} denotes the expectation with respect to ν_α) and that ν_α is stationary and satisfies (2) for the process with rates $\hat{r}^0$. For an external potential $V \in C^2(\Lambda; \mathbb{R})$ the process with rates $\hat{r}^V$ satisfies detailed balance with respect to the probability measures $\nu^V_\alpha(\eta) \propto \nu_\alpha(\eta)\mathrm{e}^{-\sum_{i\in\mathbb{T}_L^d} V(i/L)\eta(i)}$. For the measure ν^V_α, the expected number of particles at $u \in \Lambda$ is defined as

$$\bar{\rho}_{\alpha,V}(u) := \frac{E_{\nu_{\alpha,1}}\big[\eta(0)\mathrm{e}^{-V(u)\eta(0)}\big]}{E_{\nu_{\alpha,1}}\big[\mathrm{e}^{-V(u)\eta(0)}\big]} < \infty. \tag{6}$$

Combining (6) with (5) allows to show that $\bar{\rho}_{\alpha,V}(u) = (f')^{-1}(-V(u) + f'(\alpha))$, or equivalently $f'(\bar{\rho}_{\alpha,V}(u)) = -V(u) + f'(\alpha)$. Consequently (6) is strictly monotonically increasing in α. Since the number of particles is conserved, its distribution is fully determined by the initial condition for the model. In everything that follows, we restrict to initial distributions $(\mu^L_0)_{L\in\mathbb{N}}$ for which the total density of particles is bounded uniformly: there exists $C_{\mathrm{tot}} \in (0, N_{\max}]$ such that for all $L \in \mathbb{N}$

$$\mu^L_0\left(\eta \in \Omega_L \,\Big|\, \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d} \eta(i) \le C_{\mathrm{tot}}\right) = 1. \tag{7}$$

This means that the Markov chain is supported on finitely many configurations, allowing us to treat each particle system as a finite state Markov chain. Finally, for any $V \in C^2(\Lambda; \mathbb{R})$ and any α, define the relative entropy (or *microscopic free energy*) as

$$\mathcal{F}^V_{L,\alpha}(\mu) := \mathcal{H}\big(\mu|\nu^V_\alpha\big) = \sum_{\eta\in\Omega_L} \mu(\eta)\log\Big(\frac{\mu(\eta)}{\nu^V_\alpha(\eta)}\Big), \tag{8}$$

where μ is a probability measure (on Ω_L). If μ is the probability measure for our interacting particle

system at some time t then $\mathcal{F}^V_{L,\alpha}(\mu) < \infty$, by (7), since $\nu_*(\eta) > 0$ for all $\eta \in \Omega_L$.

2.1.2 Canonical Structure for Markov Chains

We now describe a Ψ-$\Psi^\star$ structure for finite state Markov chains which is related to a relative entropy between path measures [21]. This structure is central to this article (see also [29, 30]). Let μ be a probability measure on Ω_L supported on finitely many configurations. We think of this measure as a (generic) distribution of the particle system. For $\eta, \eta' \in \Omega_L$ we define the *probability current* from η to η' as

$$J_{\eta,\eta'}(\mu) := \mu(\eta) r^V_{\eta,\eta'} - \mu(\eta') r^V_{\eta',\eta}. \tag{9}$$

The divergence at η is $\operatorname{div} J(\mu)(\eta) := \sum_{\eta' \in \Omega_L} J_{\eta,\eta'}(\mu)$. Following [21], define a *mobility*

$$a_{\eta,\eta'}(\mu) := 2\big[\mu(\eta) r^V_{\eta,\eta'} \mu(\eta') r^V_{\eta',\eta}\big]^{1/2} \tag{10}$$

which is independent of V since $\hat{r}^V_{\eta,\eta'} \hat{r}^V_{\eta',\eta} = \hat{r}^0_{\eta,\eta'} \hat{r}^0_{\eta',\eta}$. Let the discrete gradient of a function h on Ω_L be $\nabla^{\eta,\eta'} h := h(\eta') - h(\eta)$ and define a *thermodynamic force* (cf. [29, 30, 21]) as

$$F^V_{\eta,\eta'}(\mu) := -\nabla^{\eta,\eta'} \log\Big(\frac{\mu}{\nu^V_\alpha}\Big), \tag{11}$$

which is in fact independent of α, as $\nu_\alpha(\eta)/\nu_\alpha(\eta^{i,i'}) = \nu_*(\eta)/\nu_*(\eta^{i,i'})$. For a general interpretation of the mobility and the force and their physical relation to thermodynamic quantities, such as entropy production and housekeeping heat, we refer the reader to [21].

The canonical structure is based on a dual paring between currents and thermodynamic forces. We consider generic currents j and forces F, which are arbitrary anti-symmetric functions on $\Omega_L \times \Omega_L$ with $j_{\eta,\eta'} = -j_{\eta',\eta}$ and $F_{\eta,\eta'} = -F_{\eta',\eta}$. The dual pairing is $\langle j, F\rangle_L := \frac{1}{2}\sum_{\eta,\eta' \in \Omega_L} j_{\eta,\eta'} F_{\eta,\eta'} \mathbf{1}_{\{a_{\eta,\eta'}(\mu) > 0\}}$ (which implicitly depends on μ). Here $\mathbf{1}_A$ is the indicator function of the event A, which is given by $\mathbf{1}_A = 1$ if the statement A is satisfied and $\mathbf{1}_A = 0$ otherwise. Now define

$$\Psi^\star_L(\mu, F) := \sum_{\eta,\eta' \in \Omega_L} a_{\eta,\eta'}(\mu)\Big[\cosh\big(\tfrac{1}{2} F_{\eta,\eta'}\big) - 1\Big] \tag{12}$$

and

$$\Psi_L(\mu, j) := \sum_{\eta,\eta' \in \Omega_L} a_{\eta,\eta'}(\mu)\Big[\frac{j_{\eta,\eta'}}{a_{\eta,\eta'}(\mu)} \operatorname{arcsinh}\Big(\frac{j_{\eta,\eta'}}{a_{\eta,\eta'}(\mu)}\Big) - \cosh\Big(\operatorname{arcsinh}\Big(\frac{j_{\eta,\eta'}}{a_{\eta,\eta'}(\mu)}\Big)\Big) + 1\Big], \tag{13}$$

where the summands in (13) have to be interpreted as being equal to zero whenever $a_{\eta,\eta'}(\mu) = 0$. The two functions (12) and (13) are both symmetric and strictly convex in their second argument. Moreover, they are Legendre dual with respect to the dual pairing $\langle j, F\rangle_L$ and give rise to the *Onsager-Machlup functional*,

$$\Phi_L(\mu, j, F) := \Psi_L(\mu, j) - \langle j, F\rangle_L + \Psi^\star_L(\mu, F) \geq 0, \tag{14}$$

where the inequality follows from the Fenchel-Young inequality (which directly follows from the Legendre duality of Ψ and $\Psi^\star$). This functional will be used in the following to characterise the relative entropy between path measures. In particular, we will study the convergence of the *non-quadratic* functionals Ψ and $\Psi^\star$ to their quadratic counterparts to a macroscopic quadratic functional, which has the form of the macroscopic Onsager-Machlup functional.

2.1.3 Projection onto the Physical Domain

So far we considered currents and densities on the full configuration space Ω_L. To obtain hydrodynamic behaviour, we 'project' the system onto the physical domain $\mathbb{T}_L^d$ and also embed the sequence of these domains (indexed by L) into the flat torus Λ. This section introduces the associated notation.

For a (generic) probability measure μ on Ω_L (which we again think of as the current distribution of the particle system), we can define the averaged number of particles $\hat{\rho}_i(\mu)$ at site $i \in \mathbb{T}_L^d$ and an averaged particle current $\hat{\jmath}^V_{i,i'}(\mu)$, as

$$\hat{\rho}_i(\mu) := \sum_{\eta\in\Omega_L} \mu(\eta)\eta(i) \qquad \text{and} \qquad \hat{\jmath}^V_{i,i'}(\mu) := \sum_{\eta\in\Omega_L} \mu(\eta)\big(\hat{r}^V_{\eta,\eta^{i,i'}} - \hat{r}^V_{\eta,\eta^{i',i}}\big). \tag{15}$$

The current $\hat{\jmath}^V_{i,i'}(\mu)$ describes the expected net flow of particles from site i to site i' if the distribution of the particle system is given by μ. For gradient dynamics and $V = 0$ the current (15) is

$$\hat{\jmath}^0_{i,i'}(\mu) = \hat{\phi}_i(\mu) - \hat{\phi}_{i'}(\mu) = -\nabla^{i,i'}\hat{\phi}(\mu), \tag{16}$$

where the discrete gradient on $\mathbb{T}_L^d$ is (for $h\colon \mathbb{T}_L^d \to \mathbb{R}$) defined as $\nabla^{i,i'}h = h(i') - h(i)$. Similar to (15), define also two (averaged) mobilities for the edge connecting i and i' as

$$\hat{a}_{i,i'}(\mu) := \sum_{\eta\in\Omega_L} 2\big[\mu(\eta)\hat{r}^V_{\eta,\eta^{i,i'}}\mu(\eta^{i,i'})\hat{r}^V_{\eta^{i,i'},\eta}\big]^{1/2}, \qquad \hat{\chi}^V_{i,i'}(\mu) := \frac{1}{2}\sum_{\eta\in\Omega_L} \mu(\eta)\big(\hat{r}^V_{\eta,\eta^{i,i'}} + \hat{r}^V_{\eta,\eta^{i',i}}\big), \tag{17}$$

which are related by $\hat{a}_{i,i'}(\mu) \le 2\hat{\chi}^V_{i,i'}(\mu)$ (with equality for $\mu = \nu^V_\alpha$). Note that the two mobilities characterise the average particle jumps between i and i' and are therefore symmetric in i and i'.

For the embedding on the flat torus, let $\mathcal{M}_+(\Lambda)$ be the set of finite and non-negative Radon measures on Λ, endowed with the weak topology. Define the empirical measure $\Theta_L\colon \Omega_L \to \mathcal{M}_+(\Lambda)$ as

$$\Theta_L(\eta) := \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d} \eta(i)\,\delta_{i/L}. \tag{18}$$

Thus, each configuration η of an interacting particle system of size L corresponds to a measure $\Theta_L(\eta) \in \mathcal{M}_+(\Lambda)$.

2.2 Path Measures on the Microscopic Scale

2.2.1 Path Measures on the Càdlàg Path Space $\mathcal{D}$

Our analysis of the hydrodynamic limit is based on the convergence of path measures. In this section, we introduce the notation that allows us to define the path measures Q_L and limit measures Q^* studied in the remainder of the article.

For any topological space $\mathcal{S}$ we denote with $\mathcal{D}([0,T];\mathcal{S})$ the set of $\mathcal{S}$ valued càdlàg paths (right-continuous paths with left limits) on $[0,T]$. For details, see [7, Chapter 3], as well as [22, Chapter 4.1] and [6]. For $t \in [0,T]$ let $X_t\colon \mathcal{D}([0,T];\mathcal{S}) \to \mathcal{S}$ be the marginal at time t, which evaluates a path $\gamma = (\gamma_t)_{t\in[0,T]} \in \mathcal{D}([0,T];\mathcal{S})$ at time t: $X_t(\gamma) = \gamma_t$. We recall that whilst X_t is measurable for all $t \in [0,T]$, it is continuous only for almost all $t \in (0,T)$, as well as $t = 0$ and $t = T$.

In the following, the expression *path measure* will refer to a probability distribution on $\mathcal{D}([0,T];\mathcal{S})$ for some $\mathcal{S}$. Let $P_L^{\tilde{V}}$ be the path measure on $\mathcal{D}([0,T];\Omega_L)$ for a particle system with time-dependent potential $\tilde{V} \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$. We can recover the distribution of this Markov chain at time t from $P_L^{\tilde{V}}$ via the push-forward measure $(X_t)_\# P_L^{\tilde{V}}$.

The hydrodynamic behaviour of the particle system depends on the behaviour of $P_L^{\tilde{V}}$ as $L \to \infty$: we use the name *reference process* for this path measure. We compare the reference process with a second path measure P_L, also on $\mathcal{D}([0,T];\Omega_L)$, which we term the *observed process*. This observed process can be any (possibly time-heterogeneous) Markov chain on Ω_L that satisfies the following properties: We assume that the associated path measure P_L on $\mathcal{D}([0,T];\Omega_L)$ is absolutely continuous with respect to $P_L^{\tilde{V}}$, that the initial condition coincides with the one of $P_L^{\tilde{V}}$, that is, $(X_0)_\# P_L = (X_0)_\# P_L^{\tilde{V}} = \mu_0^L$, and that the transition rates r_t^L are bounded in time, i.e. for each $L \in \mathbb{N}$, we assume that $\sup_{t\in[0,T]}(r_t^L)_{\eta,\eta'} < \infty$ for all $\eta, \eta' \in \Omega_L$.

We can assign to P_L a unique path $(\mu_t^L, \jmath_t^L)_{t\in[0,T]}$ consisting of the density $\mu_t^L := (X_t)_\# P_L$ and the current $(\jmath_t^L)_{\eta,\eta'} := \mu_t^L(\eta)(r_t^L)_{\eta,\eta'} - \mu_t^L(\eta')(r_t^L)_{\eta',\eta}$, which are again linked by a continuity equation $\partial_t \mu_t^L = -\operatorname{div} \jmath_t^L$.

We remark that for the choice $P_L = P_L^{\tilde{V}}$ the current $\jmath_t^L$ simply coincides with the probability current (9) for the time-dependent rate $r^{\tilde{V}_t}$. In this case, one can further show that the associated density and current (15) satisfy the continuity equation $\partial_t \hat{\rho}_i(\mu_t^L) = -\operatorname{div} \hat{\jmath}^V(\mu_t^L)(i)$, where the divergence on the physical domain $\mathbb{T}_L^d$ is defined as $\operatorname{div} \hat{\jmath}^V(\mu)(i) := \sum_{i'\in\mathbb{T}_L^d} \hat{\jmath}^V_{i,i'}(\mu)$.

Since every Ω_L can be embedded into the flat torus Λ (as a map from Ω_L to $\mathcal{M}_L(\Lambda)$), there is a corresponding embedding of the path space $\mathcal{D}([0,T];\Omega_L)$ into $\mathcal{D}([0,T];\mathcal{M}_L(\Lambda))$. In particular, each path measure Q_L on $\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$ that is supported on $\mathcal{M}_L(\Lambda) := \{L^{-d}\sum_{i\in\mathbb{T}_L^d} k_i \delta_{i/L} \,|\, k_i \in \mathbb{N}_0, k_i \le N_{\max}\}$ can be identified with a unique measure P_L on $\mathcal{D}([0,T];\Omega_L)$. The measure on $\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$ that corresponds to the reference process $P_L^{\tilde{V}}$ is denoted with $Q_L^{\tilde{V}}$. Similarly, for the observed process, there is a Q_L corresponding to P_L. No information is lost on embedding the processes into Λ, so $\mathcal{H}(Q_L|Q_L^{\tilde{V}}) = \mathcal{H}(P_L|P_L^{\tilde{V}})$, which can be proved by two applications of Lemma 9.4.5 in [1] with the bijection from $\mathcal{M}_L(\Lambda)$ to Ω_L.

2.2.2 Microscopic Action Functional

To compare the reference and the observed process, consider the thermodynamic force for the reference process at time t, which is $F^{\tilde{V}_t}(\mu_t^L)$, evaluated from (11) with $\mu_t^L = (X_t)_\# P_L$. Since P_L is absolutely continuous with respect to $P_L^{\tilde{V}}$, the relative entropy $\mathcal{H}(P_L|P_L^{\tilde{V}})$ is under the assumptions in Section 2.2.1 finite and (cf. [21, Appendix]) coincides with

$$\mathcal{H}\big(P_L|P_L^{\tilde{V}}\big) = \mathcal{H}\big(\mu_0^L|(X_0)_\# P_L^{\tilde{V}}\big) + \frac{1}{2}\int_0^T \Phi_L\big(\mu_t^L, \jmath_t^L, F^{\tilde{V}_t}(\mu_t^L)\big)\,\mathrm{d}t. \tag{19}$$

Moreover, $\mathcal{H}(\mu_0^L|(X_0)_\# P_L^{\tilde{V}}) = 0$, since P_L and $P_L^{\tilde{V}}$ share the same initial condition. We interpret $\frac{1}{2}\Phi_L(\mu_t^L, \jmath_t^L, F_\alpha^{\tilde{V}_t}(\mu_t^L))$ as an extended Lagrangian [21] and define the *microscopic action* of the path measure Q_L as the relative entropy

$$\mathbb{A}_L^{\tilde{V}}\big(Q_L\big) := \mathcal{H}\big(Q_L|Q_L^{\tilde{V}}\big) = \mathcal{H}\big(P_L|P_L^{\tilde{V}}\big) = \frac{1}{2}\int_0^T \Phi_L\big(\mu_t^L, \jmath_t^L, F^{\tilde{V}_t}(\mu_t^L)\big)\,\mathrm{d}t. \tag{20}$$

This is the central functional defined on the discrete (lattice) level studied in this article.

2.3 Macroscopic Quantities

In the hydrodynamic scaling limit, the microscopic action (20) will converge to a macroscopic action, which is (30). (For the macroscopic setting, we restrict our considerations to potentials V that are constant in time.) We now show how the macroscopic action functional is constructed.

2.3.1 The Macroscopic Free Energy

For $\alpha \in (0, N_{\max}]$ and $V \in C^2(\Lambda;\mathbb{R})$, we define the *macroscopic free energy* $\mathcal{F}_\alpha^V : \mathcal{M}_+(\Lambda) \to [0,\infty]$ as

$$\mathcal{F}_\alpha^V(\pi) := \sup_{h\in C(\Lambda;\mathbb{R})} \left[\langle \pi, h\rangle - \int_\Lambda \log\left(\frac{Z_1(f'(a)+h(u)-V(u))}{Z_1(f'(a)-V(u))} \right) \mathrm{d}u \right]. \tag{21}$$

This free energy coincides with a rate function: there is a large-deviation principle for the particle configuration Θ_L sampled from the the steady state ν_α^V; the speed of this LDP is L^d and its rate function is $\mathcal{F}_\alpha^V(\pi)$, (see e.g. Section 5.1, page 75 in [22] for the special case of a zero-range process). From (3), $\mathcal{F}_\alpha^V(\pi)$ is finite only if $\pi(\mathrm{d}u) = \rho(u)\mathrm{d}u$ for some density $\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$. In the following we thus write $\mathcal{F}_\alpha^V(\rho)$ for $\mathcal{F}_\alpha^V(\pi)$. As in Macroscopic Fluctuation Theory [5, Section 5.A], we can represent $\mathcal{F}_\alpha^V$ for reversible systems as

$$\mathcal{F}_\alpha^V(\rho) = \int_\Lambda \Big[f(\rho(u)) - f(\bar\rho_{\alpha,V}(u)) - f'(\bar\rho_{\alpha,V}(u))\big(\rho(u) - \bar\rho_{\alpha,V}(u)\big) \Big] \mathrm{d}u, \tag{22}$$

where $\bar\rho_{\alpha,V} \in \mathcal{L}^1(\Lambda;[0,\infty))$, introduced in (6), is the steady state density for the dynamics of the macroscopic system. Note that (22) inherits the convexity of f.

2.3.2 The Hydrodynamic Current and the Hydrodynamic Equation

In the hydrodynamic limit, the particle density at time t is given by some $\rho_t \in \mathcal{L}^1(\Lambda;[0,\infty))$. The hydrodynamic current describes the resulting particle flow:

$$J(\rho) := -\nabla\phi(\rho) - \chi(\rho)\nabla V, \tag{23}$$

where ϕ and χ are functions that depend on the system of interest and are discussed later in this section. The hydrodynamic equation is then

$$\dot\rho_t = -\nabla\cdot J(\rho_t) = \Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V). \tag{24}$$

In this article, we consider weak solutions to (24), in the sense that for all $G \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$

$$\begin{aligned} \int_\Lambda \rho_T G_T \,\mathrm{d}u - \int_\Lambda \rho_0 G_0 \,\mathrm{d}u - \int_0^T \int_\Lambda \rho_t\, \partial_t G_t \,\mathrm{d}u\,\mathrm{d}t \\ = \int_0^T \int_\Lambda \phi(\rho_t)\Delta G_t \,\mathrm{d}u\,\mathrm{d}t - \int_0^T \int_\Lambda \chi(\rho_t)\nabla V\cdot\nabla G_t \,\mathrm{d}u\,\mathrm{d}t. \end{aligned} \tag{25}$$

The dynamics on the macroscopic scale are characterised by the functions ϕ, χ in (24). To relate these quantities to the microscopic dynamics, we consider the case $V = 0$, so that $E_{\nu_{\alpha,1}}[\eta(0)] = \alpha$. Define the macroscopic mobility $\chi\colon [0, N_{\max}] \to [0,\infty)$ as

$$\chi(\alpha) := \hat\chi^0_{i,i+e_k}(\nu_\alpha) = \frac{1}{2}\hat a_{i,i+e_k}(\nu_\alpha), \tag{26}$$

which is independent of i and e_k (and thus well-defined). To see this, note from (2) and (17) that $\hat\chi^0_{i,i+e_k}(\nu_\alpha) = \sum_{\eta\in\Omega_L} \nu_\alpha(\eta)\hat r^0_{\eta,\eta^{i,i+e_k}} = E_{\nu_{\alpha,1}}[g_1(\eta(0))]E_{\nu_{\alpha,1}}[g_2(\eta(0))]$, where we used (1) and the product structure of ν_α. Similarly, define $\phi\colon [0,N_{\max}] \to [0,\infty)$ by $\phi(\alpha) := \hat\phi_i(\nu_\alpha) = E_{\nu_{\alpha,1}}[\mathrm{d}(\eta(0))]$, which is by construction independent of i. One then can prove the *local Einstein relation*

$$\phi'(\alpha) = f''(\alpha)\chi(\alpha), \tag{27}$$

which relates ϕ and χ to the free energy f from Section 2.3.1. Equation (27) can be obtained by differentiating $\phi(\alpha) = E_{\nu_{*,1}}[\mathrm{d}(\eta(0))\mathrm{e}^{f'(\alpha)\eta(0)}]/E_{\nu_{*,1}}[\mathrm{e}^{f'(\alpha)\eta(0)}]$. Note that $\phi'(\alpha) = \frac{1}{2}f''(\alpha)\sum_\eta \nu_\alpha(\eta)\big[\mathrm{d}(\eta(i)) - \mathrm{d}(\eta(i'))\big](\eta(i)-\eta(i'))$ (for $i, i' \in \mathbb{T}_L^d$ arbitrary with $i \neq i'$). Further, the gradient structure and detailed balance yield $\frac{1}{2}\sum_\eta \nu_\alpha(\eta)\big[\hat{r}^0_{\eta,\eta^{i,i'}} - \hat{r}^0_{\eta,\eta^{i',i}}\big](\eta(i)-\eta(i')) = \frac{1}{2}\sum_\eta \nu_\alpha(\eta)\big[\hat{r}^0_{\eta,\eta^{i,i'}} + \hat{r}^0_{\eta,\eta^{i',i}}\big] = \chi(\alpha)$.

2.3.3 The Macroscopic Action Functional and the Chain Rule

For $\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ and $h\colon \Lambda \to \mathbb{R}^d$, we introduce the norm $\|h\|^2_{\chi(\rho)} := \int_\Lambda \chi(\rho(u))|h(u)|^2\,\mathrm{d}u$ (for full details and associated spaces, see Section 4 below). The macroscopic analogues of the (time integrals of the) microscopic functions Ψ_L and $\Psi_L^\star$ from (12), (13) are

$$\mathcal{E}\big((\rho_t)_{t\in[0,T]}\big) := \sup_G \bigg[\bigg(\int_\Lambda \rho_T G_T\,\mathrm{d}u - \int_\Lambda \rho_0 G_0\,\mathrm{d}u - \int_0^T\int_\Lambda \rho_t\,\partial_t G_t\,\mathrm{d}u\,\mathrm{d}t\bigg) - \frac{1}{2}\int_0^T \|\nabla G_t\|^2_{\chi(\rho_t)}\,\mathrm{d}t\bigg] \tag{28}$$

and

$$\mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big) := \sup_G \bigg[\bigg(\int_0^T\int_\Lambda \phi(\rho_t)\Delta G_t\,\mathrm{d}u\,\mathrm{d}t - \int_0^T\int_\Lambda \chi(\rho_t)\nabla V\cdot\nabla G_t\,\mathrm{d}u\,\mathrm{d}t\bigg) - \frac{1}{2}\int_0^T \|\nabla G_t\|^2_{\chi(\rho_t)}\,\mathrm{d}t\bigg], \tag{29}$$

where the supremum is in both cases over $C^{1,2}([0,T]\times\Lambda;\mathbb{R})$. We will show in Propositions 4.1 and 4.4 that, under certain assumptions, these functionals can be expressed as time integrals of suitably defined norms

$$\mathcal{E}\big((\rho_t)_{t\in[0,T]}\big) = \frac{1}{2}\int_0^T \|\dot\rho_t\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t$$

and

$$\mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big) = \frac{1}{2}\int_0^T \|\Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t = \frac{1}{2}\int_0^T \|f''(\rho_t)\nabla\rho_t + \nabla V\|^2_{\chi(\rho_t)}\,\mathrm{d}t.$$

In particular, we will show that non-quadratic Ψ and $\Psi^\star$ of (13) and (12) can be bounded by the quadratic expressions $\mathcal{E}$ and $\mathcal{E}^\star$, respectively.

Finally, for $(\pi_t)_{t\in[0,T]}$ absolutely continuous with respect to the Lebesgue measure, we define the *macroscopic action* as

$$\mathbb{A}\big((\pi_t)_{t\in[0,T]}\big) := \frac{1}{2}\big[\mathcal{F}^V_\alpha(\rho_T) - \mathcal{F}^V_\alpha(\rho_0) + \mathcal{E}\big((\rho_t)_{t\in[0,T]}\big) + \mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big)\big]. \tag{30}$$

If $(\pi_t)_{t\in[0,T]}$ is not absolutely continuous with respect to the Lebesgue measure, we set $\mathbb{A}\big((\pi_t)_{t\in[0,T]}\big) = +\infty$.

In a nutshell, the main results of this article are twofold: Firstly, we establish relations between suitably scaled $\mathbb{A}_L^{\tilde V}$ of (20) and the continuum limit (30): see Theorems 3.4 to 3.6. Secondly, we show that under suitable regularity assumptions, in particular if the free energy $\mathcal{F}^V_\alpha$ satisfies a chain rule (see Equation (39)), the macroscopic action can be re-written in a way which reveals the hydrodynamic limit as minimiser of this functional, see (40) below.

2.4 Assumptions on the Particle Systems Studied

2.4.1 Local Equilibrium Assumption and the Replacement Lemma

When taking the hydrodynamic limit, one must prove a local equilibration condition, which means that the system resembles — in a small neighbourhood around any point — an equilibrium system. To make this precise, take $\ell \in \mathbb{N}$ and define the average number of particles in a box with diameter

$2\ell+1$ as

$$\eta^\ell(i) := \frac{1}{(2\ell+1)^d} \sum_{|m|\le\ell} \eta(i+m).$$

Similarly, we also define the averages $\hat\chi^\ell_{i,i+e_k}(\mu) := (2\ell+1)^{-d}\sum_{|m|\le\ell}\hat\chi_{i+m,i+m+e_k}(\mu)$ and $\hat\phi^\ell_i(\mu) := (2\ell+1)^{-d}\sum_{|m|\le\ell}\hat\phi_{i+m}(\mu)$.

Now assume that $L \gg 1$ and $\epsilon \ll 1$ and that the state of the system is given by $\eta \in \Omega_L$. Define $\ell = \lfloor \epsilon L \rfloor$, which is the size of a macroscopic box with diameter $\approx 2\epsilon$ (measured on the macroscopic scale). Hence $\hat\chi^{\lfloor\epsilon L\rfloor}_{i,i+e_k}(\delta_\eta)$ is a locally averaged mobility. *Local equilibration* means that $\hat\chi_{i,i+e_k}(\nu_{\eta^{\lfloor\epsilon L\rfloor}(i)})$ is close to the expected mobility for an equilibrium distribution ν_α with the same (locally-averaged) particle density. That is, the time averaged distributions $\mu^L_{[0,T]} := \frac{1}{T}\int_0^T \mu^L_t \,\mathrm{d}t$ satisfy in local equilibrium

$$\limsup_{\epsilon\to0}\limsup_{L\to\infty}\frac{1}{L^d}\sum_{i\in\mathbb{T}^d_L}\sum_{k=1}^d\sum_{\eta\in\Omega_L}\mu^L_{[0,T]}(\eta)\left|\hat\chi^{\lfloor\epsilon L\rfloor}_{i,i+e_k}(\delta_\eta)-\hat\chi_{i,i+e_k}(\nu_{\eta^{\lfloor\epsilon L\rfloor}(i)})\right| = 0, \tag{31}$$

as well as

$$\limsup_{\epsilon\to0}\limsup_{L\to\infty}\frac{1}{L^d}\sum_{i\in\mathbb{T}^d_L}\sum_{\eta\in\Omega_L}\mu^L_{[0,T]}(\eta)\left|\hat\phi^{\lfloor\epsilon L\rfloor}_{i}(\delta_\eta)-\hat\phi_{i}(\nu_{\eta^{\lfloor\epsilon L\rfloor}(i)})\right| = 0. \tag{32}$$

Remark (Replacement Lemma). Note that results like (31) and (32) are classically obtained by proving the stronger replacement lemma, which in our notation amounts to proving for $\hat\chi$ (and analogously for $\hat\phi$)

$$\limsup_{\epsilon\to0}\limsup_{L\to\infty}\sup_{\mu}\frac{1}{L^d}\sum_{i\in\mathbb{T}^d_L}\sum_{k=1}^d\sum_{\eta\in\Omega_L}\mu(\eta)\left|\hat\chi^{\lfloor\epsilon L\rfloor}_{i,i+e_k}(\delta_\eta)-\hat\chi_{i,i+e_k}(\nu_{\eta^{\lfloor\epsilon L\rfloor}(i)})\right| = 0, \tag{33}$$

where the supremum is taken over a class of measures μ satisfying certain bounds on the relative entropy (i.e. the free energy) and the Dirichlet form, which can be identified with $\frac{1}{2}\Psi^\star(\mu, F^V(\mu))$ (see e.g. the remark in the proof of Proposition 5.4 below). In the following, we will follow the classical approach and work with (33). We state sufficient conditions for the replacement lemma in Section 3.2 below and establish in this way the validity of (31) and (32).

2.4.2 Assumptions on the Path Measures $P_L^{\tilde V}$

We have presented a general framework for interacting particles on lattices and their hydrodynamic scaling limits. The results of the next section are similarly general and can be applied to a range of systems, including the symmetric simple exclusion process and certain zero-range processes, as discussed in Section 3.4 below. However, our results for hydrodynamic limits clearly do not apply to all interacting-particle systems. We summarise here the main assumptions on the reference process $P_L^{\tilde V}$ required in the following analysis: these need to be verified in order to apply our results to a particular system.

On the microscopic scale, we assume that the transition rates are given by (1) and are of gradient type. The initial conditions and invariant measures are as described in Section 2.1.1. We note that many of the proofs given below make use of assumption (7). Despite the fact that it is a non-standard assumption for hydrodynamic limits (unless $N_{\max} < \infty$, in which case (7) holds trivially), it is not too restrictive, in the sense that the typical initial conditions $(\mu^L_0)_{L\in\mathbb{N}}$ can be shown to satisfy (cf. equation (1.4) in Section 5.1 on page 71 in [22]) $\lim_{A\to\infty}\limsup_{L\to\infty}\mu^L_0(\eta\in\Omega_L \,|\, L^{-d}\sum_{i\in\mathbb{T}^d_L}\eta(i)\ge A) = 0$.

When taking the hydrodynamic limit, we assume that for any sequence of measures $(\mu^L)_{L\in\mathbb{N}}$ satisfying (7), it holds that

$$C_{\hat{\chi}} := \limsup_{L\to\infty} \frac{1}{L^d} \sum_{i\in\mathbb{T}_L^d} \sum_{k=1}^{d} \hat{\chi}_{i,i+e_k}(\mu^L) < \infty, \tag{34}$$

which ensures that the total rate of particle jumps for the reference process stays controlled as $L \to \infty$. Similarly we suppose that any sequence of measures $(\mu^L)_{L\in\mathbb{N}}$ obeying (7) also satisfies

$$C_{\hat{\phi}} := \limsup_{L\to\infty} \frac{1}{L^d} \sum_{i\in\mathbb{T}_L^d} \hat{\phi}_i(\mu^L) < \infty. \tag{35}$$

In addition, our proofs require the following technical assumptions on the functions f, ϕ and χ that characterise the hydrodynamic limit itself: We assume that $f \in C^2([0, N_{\max}]; \mathbb{R})$ with $f(0) = 0$, $f'' > 0$ a.e. and that $\lim_{r\to 0} f'(r) = -\infty$ and $\lim_{r\to N_{\max}} f'(r) = \infty$. Note that this implies by (5) that $\phi(0) = 0 = \chi(0)$. Further, we assume that $\phi, \chi > 0$ on $(0, N_{\max})$ and that both ϕ and χ are Lipschitz continuous on $[0, N_{\max}]$, without loss of generality with common Lipschitz constant $C_{\mathrm{Lip}} > 0$. Since $\phi(0) = \chi(0) = 0$, we have in particular $0 < \phi(a), \chi(a) \le C_{\mathrm{Lip}} a$ for $a \in (0, N_{\max}]$. We further assume that ϕ is continuously differentiable on $(0, N_{\max})$ (by the above Lipschitz condition with bounded derivative) and also strictly monotonically increasing. This implies the existence of a continuous inverse $\phi^{-1}\colon \phi([0, N_{\max}]) \to [0, N_{\max}]$, where $\phi([0, N_{\max}]) = \{\phi(a) : a \in [0, N_{\max}]\}$. We also suppose that ϕ^{-1} has a bounded derivative (which is by the inverse function theorem equivalent to saying that there exists $C_* > 0$ such that $\phi'(a) \ge C_*$ for all $a \in (0, N_{\max}]$).

3 Statement of the Results

In this section, we discuss the behaviour of the microscopic action in the limit $L \to \infty$, and the implications of this behaviour for hydrodynamic limits. Sections 3.1 and 3.2 derive preliminary results, which establish properties of the action functionals and sufficient conditions for local equilibration. Section 3.3 states the main results, consisting of three theorems (Theorems 3.4–3.6). Finally Section 3.4 discusses the applications of these theorems in two specific particle systems, and their implications for hydrodynamic limits.

3.1 Properties of the Microscopic and Macroscopic Action Functions

3.1.1 Chain rule on Microscopic Scale

Consider $(\mu_t^L, \jmath_t^L)_{t\in[0,T]}$ as in Section 2.2.1. The force $F^V(\mu_t^L)$ can be linked to the free energy (8) via the classical chain rule formula (cf. Theorem 9.2 of Appendix 1 in [22], Proposition 2.2 in [18] and also [21]) $\mathcal{F}^V_{L,\alpha}(\mu^L_{t_2}) - \mathcal{F}^V_{L,\alpha}(\mu^L_{t_1}) = -\int_{t_1}^{t_2} \langle \jmath_t^L, F^V(\mu_t^L)\rangle_L \,\mathrm{d}t$, which is a special case of the following result (proved in Section 5.1 below).

Proposition 3.1 (Chain rule for the microscopic free energy)**.** *Let $\tilde{V} \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ and consider a path measure P_L on Ω_L, as described in Section 2.2.1, with associated density and current $(\mu_t^L, \jmath_t^L)_{t\in[0,T]}$. Then the map $t \mapsto \mathcal{F}^{\tilde{V}_t}_{L,\alpha}(\mu_t^L)$ is absolutely continuous for $t \in [0,T]$ and satisfies the following chain rule. For all $0 \le t_1 < t_2 \le T$*

$$\mathcal{F}^{\tilde{V}_{t_2}}_{L,\alpha}(\mu^L_{t_2}) - \mathcal{F}^{\tilde{V}_{t_1}}_{L,\alpha}(\mu^L_{t_1}) = -\int_{t_1}^{t_2} \langle \jmath_t^L, F^{\tilde{V}_t}(\mu_t^L)\rangle_L \,\mathrm{d}t + \int_{t_1}^{t_2} \sum_{i\in\mathbb{T}_L^d} \big(\hat{\rho}_i(\mu_t^L) - \bar{\rho}_{\alpha,\tilde{V}_t}(i)\big)\, \partial_t \tilde{V}_t(\tfrac{i}{L})\,\mathrm{d}t. \tag{36}$$

Now fix some $\alpha \in (0, N_{\max})$ and combine Proposition 3.1 with (14) and (19), which yields

$$\mathbb{A}_L^{\tilde{V}}(Q_L) = \frac{1}{2}\big[\mathcal{F}_{L,\alpha}^{\tilde{V}_T}(\mu_T^L) - \mathcal{F}_{L,\alpha}^{\tilde{V}_0}(\mu_0^L)\big] + \frac{1}{2}\int_0^T \Psi_L(\mu_t^L, \jmath_t^L)\,\mathrm{d}t + \frac{1}{2}\int_0^T \Psi_L^\star\big(\mu_t^L, F^{\tilde{V}_t}(\mu_t^L)\big)\,\mathrm{d}t \\ - \frac{1}{2}\int_0^T \sum_{i\in\mathbb{T}_L^d} \big(\hat{\rho}_i(\mu_t^L) - \bar{\rho}_{\alpha,\tilde{V}_t}(i)\big)\,\partial_t \tilde{V}_t(\tfrac{i}{L})\,\mathrm{d}t \geq 0. \quad (37)$$

3.1.2 Macroscopic Action

We now establish some properties of $\mathbb{A}$, as defined in (30). If $\mathbb{A}\big((\pi_t)_{t\in[0,T]}\big) < \infty$ one can show that

$$\mathbb{A}\big((\pi_t)_{t\in[0,T]}\big) = \frac{1}{2}\big[\mathcal{F}_\alpha^V(\rho_T) - \mathcal{F}_\alpha^V(\rho_0)\big] + \frac{1}{4}\int_0^T \big(\|\dot{\rho}_t\|^2_{-1,\chi(\rho_t)} + \|\Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}\big)\,\mathrm{d}t, \quad (38)$$

see Proposition 4.1 and Proposition 4.4. For a definition of the norm $\|\cdot\|_{-1,\chi(\rho_t)}$ (and the associated inner product $\langle\cdot,\cdot\rangle_{-1,\chi(\rho_t)}$) we also refer to Section 4 below.

Note that $\mathbb{A}((\pi_t)_{t\in[0,T]})$ as defined here might in general be negative. A sufficient condition for non-negativity of $\mathbb{A}((\pi_t)_{t\in[0,T]})$ is ensured by the validity of the following chain rule, which can be seen as a macroscopic counterpart to (36) for potentials constant in time. A *formal* calculation yields for $0 \leq t_1 < t_2 \leq T$ the chain rule

$$\mathcal{F}_\alpha^V(\rho_{t_2}) - \mathcal{F}_\alpha^V(\rho_{t_1}) = \int_{t_1}^{t_2} \big\langle \dot{\rho}_t, \frac{\delta \mathcal{F}_\alpha^V}{\delta \rho_t}\big\rangle\,\mathrm{d}t = -\int_{t_1}^{t_2} \langle \dot{\rho}_t, \Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\rangle_{-1,\chi(\rho_t)}\,\mathrm{d}t. \quad (39)$$

Combined with (38) this allows us to (formally!) rewrite the macroscopic action functional (38) as

$$\mathbb{A}\big((\pi_t)_{t\in[0,T]}\big) = \frac{1}{4}\int_0^T \big\|\dot{\rho}_t - \Delta\phi(\rho_t) - \nabla\cdot(\chi(\rho_t)\nabla V)\big\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t. \quad (40)$$

In Section 4.2 we summarise some geometrical properties of the relevant function spaces and we establish sufficient conditions for the chain rule:

Theorem 3.2. *Let the assumptions from Section 2.4.2 hold and additionally assume that $\chi'(a) \geq C_*$ for all $a \in (0, N_{\max}]$ (for some $C_* > 0$). If $d > 1$, then further assume that the free energy density f satisfies the McCann condition for geodesic convexity (stated in Equation (70) below). Then any path $(\pi_t)_{t\in[0,T]}$ with $\mathbb{A}((\pi_t)_{t\in[0,T]}) < \infty$ and $\mathcal{F}_\alpha^V(\rho_0) < \infty$ satisfies the identities in Equation (39).*

Note that the McCann condition is always satisfied in one spatial dimension (where it reduces to convexity of f). We further stress that in Macroscopic Fluctuation Theory the validity of the chain rule is implicitly assumed by Equation (2.15) in [5], which relates the large deviation rate for a forward path to its time-reversed counterpart.

3.2 Sufficient Conditions for Local Equilibration

The following theorem, proved in Section 5.1 below, yields a sufficient condition for the local equilibration discussed in Section 2.4.1 in terms of the free energy (8) of the initial condition and the action functional (20).

Theorem 3.3. *Let $(P_L)_{L\in\mathbb{N}}$ be as in Section 2.2.1 with densities $(\mu_t^L)_{t\in[0,T]}$, for $L \in \mathbb{N}$, and associated path measures $(Q_L)_{L\in\mathbb{N}}$ on $\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$. Assume there exist $V \in C^2(\Lambda;\mathbb{R})$ and $\alpha \in [0, N_{\max})$ such that*

$$\limsup_{L\to\infty} \frac{1}{L^d}\mathcal{F}_{L,\alpha}^V(\mu_0^L) < \infty \quad (41)$$

and $\tilde{V} \in C^{1,2}([0,T] \times \Lambda; \mathbb{R})$ *such that*

$$\limsup_{L\to\infty} \frac{1}{L^d} \mathbb{A}_L^{\tilde{V}}(Q_L) < \infty. \tag{42}$$

Then $(\mu^L_{[0,T]})_{L\in\mathbb{N}}$ *satisfies the local equilibrium assumption,* (31) *and* (32)*. Moreover, Equations* (41) *and* (42) *are independent of* V*,* $\tilde{V}$ *and* α*, such that these conditions can equivalently be stated as* $\limsup_{L\to\infty} L^{-d}\mathcal{H}(Q_L|Q_{\nu_\alpha}) < \infty$*, where* Q_{ν_α} *denotes the measure on* $\mathcal{D}([0,T];\Omega_L)$ *with marginals equal to* ν_α*, in the sense that* $(X_t)_\# Q_{\nu_\alpha} = (\Theta_L)_\# \nu_\alpha$ *for all* $t \in [0,T]$*.*

3.3 Particle Systems on Hydrodynamic Scale

We now present our main results. We consider sequences of path measures $(Q_L^V)_{L\in\mathbb{N}}$ and $(Q_L)_{L\in\mathbb{N}}$ on $\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$, as defined in Section 2.2.1, as well as the corresponding sequences $(P_L^V)_{L\in\mathbb{N}}$ and $(P_L)_{L\in\mathbb{N}}$. We define Q^* as a (possibly non-unique) limit point of the sequence of observed processes $(Q_L)_{L\in\mathbb{N}}$ and we establish various properties of this limit. The physical idea is that the path on which Q^* is supported is a *candidate* for the hydrodynamic limit for the reference process $(Q_L^V)_{L\in\mathbb{N}}$. By analysing the large-L behaviour of the microscopic action $\mathbb{A}_L^V(Q_L)$, the aim is to show that the only admissible candidate path is the true hydrodynamic limit. For specific examples, see Section 3.4, below.

3.3.1 Assumptions for Scaling Limits

To apply the results of this section to a specific interacting particle system (reference process), several assumptions have to be satisfied. We assume that the conditions given in Section 2.4.2 have been verified. We assume also that the initial distributions $(\mu_0^L)_{L\in\mathbb{N}}$ of $(P_L^V)_{L\in\mathbb{N}}$ converge to a fixed density $\rho_0 \in \mathcal{L}^1(\Lambda;[0,\infty))$ in the sense that $(\Theta_L)_\#\mu_0^L \to \delta_{\pi_0}$ with $\pi_0(\mathrm{d}u) = \rho_0(u)\mathrm{d}u$. For the rest of this Section 3.3, we fix α uniquely by requiring that $\int_\Lambda \rho_0(u)\,\mathrm{d}u = \int_\Lambda \bar{\rho}_{\alpha,V}(u)\,\mathrm{d}u$.

Further, we assume that the $(Q_L)_{L\in\mathbb{N}}$ are relatively compact [7, 22]. Then there is a measure Q^* on $\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$ and a subsequence of $(Q_L)_{L\in\mathbb{N}}$ converging to Q^* (such that the marginal at time $t = 0$ satisfies $(X_0)_\# Q^* = \delta_{\pi_0}$). Finally, we assume that the measure Q^* is concentrated on paths that are absolutely continuous with respect to the Lebesgue measure,

$$Q^*\Big((\pi_t)_{t\in[0,T]} \in \mathcal{D}([0,T];\mathcal{M}_+(\Lambda)) : \pi_t(\mathrm{d}u) = \rho_t(u)\,\mathrm{d}u \text{ for a.a. } t \in [0,T]\Big) = 1. \tag{43}$$

We note that the paths in (43) satisfy $\rho_t \in \mathcal{L}^1(\Lambda;[0,\infty))$. Moreover, if $N_{\max} < \infty$, then clearly also $\rho_t \le N_{\max}$ a.e. on Λ for almost all $t \in [0,T]$. However, the limit Q^* is not assumed to be unique: there could exist other subsequences of $(Q_L)_{L\in\mathbb{N}}$ with different limits.

Given a specific model, the compactness of the sequence $(Q_L)_{L\in\mathbb{N}}$ and the support on absolutely continuous paths (43) often follow from (41) in combination with an assumptions on the transition rates of the particle system. This is the case for the examples considered in Section 3.4 below.

3.3.2 Comparison with classical proofs of the Hydrodynamic Limit

To provide context for our analysis, we briefly summarise the classical approach to hydrodynamic limits. Here, we consider separately the observed process and the reference process, but the classical approach takes $(P_L)_{L\in\mathbb{N}} = (P_L^V)_{L\in\mathbb{N}}$. The task of proving a hydrodynamic limit for $(Q_L)_{L\in\mathbb{N}}$ then consists of characterising all limiting distributions. The first step is to establish relative compactness [7, 22], which ensures the existence of a (possibly non-unique) limit Q^*. One then shows that Q^* is unique and that it is concentrated on a single path $(\rho_t)_{t\in[0,T]}$ (i.e. $Q^* = \delta_{(\pi_t)_{t\in[0,T]}}$ and $\pi_t(\mathrm{d}u) = \rho_t(u)\mathrm{d}u$

for almost all $t \in [0,T]$). This general approach includes both the entropy method and the relative entropy method [22]: note that it *first* establishes that Q^* is supported on weak solutions to (25) and *then* uses a uniqueness result for this solution to infer that Q^* is supported on this unique solution, see e.g. [22, Chapter 4].

Our approach here differs in two main points: We consider an observed process that is different from the reference process ($P_L \neq P_L^V$ in general) and we assume that the sequence $(Q_L)_L$ has a unique limiting distribution Q^* that is concentrated on a single path. (As a special case, one may take $P_L = P_L^V$, under the assumption that the hydrodynamic limit exists, but the following results are not restricted to this case.)

3.3.3 Convergence of Free Energy and Action for Deterministic Limits

The following first main theorem yields regularity results for $(P_L)_{L\in\mathbb{N}}$ under the assumptions of Section 3.3.1 and those of Theorem 3.3. In particular, it shows that the macroscopic action (and its individual contributions) are asymptotically dominated by their (more detailed) microscopic counterparts.

Theorem 3.4 (Regularity of the limit and asymptotic lower bounds)**.** *Let $(P_L)_{L\in\mathbb{N}}$ be a sequence as in Section 3.3.1, with density and current $(\mu_t^L, \jmath_t^L)_{t\in[0,T]}$, for $L \in \mathbb{N}$. We suppose that the associated sequence $(Q_L)_{L\in\mathbb{N}}$ has a unique limit point $Q^* = \delta_{(\pi_t)_{t\in[0,T]}}$ for some $(\pi_t)_{t\in[0,T]} \in \mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$ and that the initial condition is well prepared in the sense that the free energies converge (cf. [37, 18, 33])*

$$\lim_{L\to\infty} \frac{1}{L^d} \mathcal{F}_{L,\alpha}^V(\mu_0^L) = \mathcal{F}_\alpha^V(\rho_0). \tag{44}$$

Further assume that $(Q_L)_{L\in\mathbb{N}}$ satisfies (42) for $\tilde{V}_t = V$, such that

$$\limsup_{L\to\infty} \frac{1}{L^d} \mathbb{A}_L^V(Q_L) < \infty. \tag{45}$$

Then $(\pi_t)_{t\in[0,T]}$ is narrowly continuous, i.e. $(\pi_t)_{t\in[0,T]} \in C([0,T];\mathcal{M}_+(\Lambda))$ and the action satisfies the lower bound

$$\liminf_{L\to\infty} \frac{1}{L^d} \mathbb{A}_L^V(Q_L) \geq \mathbb{A}((\pi_t)_{t\in[0,T]}). \tag{46}$$

Further, the free energy satisfies for all $t \in [0,T]$

$$\liminf_{L\to\infty} \frac{1}{L^d} \mathcal{F}_{L,\alpha}^V(\mu_t^L) \geq \mathcal{F}_\alpha^V(\rho_t), \tag{47}$$

as well as

$$\liminf_{L\to\infty} \frac{1}{L^d} \int_0^T \Psi_L(\mu_t^L, \jmath_t^L)\,\mathrm{d}t \geq \frac{1}{2}\int_0^T \|\dot{\rho}_t\|_{-1,\chi(\rho_t)}^2\,\mathrm{d}t \tag{48}$$

and

$$\liminf_{L\to\infty} \frac{1}{L^d} \int_0^T \Psi_L^\star\big(\mu_t^L, F^V(\mu_t^L)\big)\,\mathrm{d}t \geq \frac{1}{2}\int_0^T \|\Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\|_{-1,\chi(\rho_t)}^2\,\mathrm{d}t. \tag{49}$$

In this theorem, we see for the first time a connection between the non-quadratic microscopic functionals Ψ and $\Psi^\star$ and their macroscopic quadratic counterparts, see (48) and (49).

Proof. Note that the assumptions of Theorem 3.3 are satisfied, so that the local equilibration assumptions (31) and (32) hold. The result (46) follows from the representation of $\mathbb{A}_L^V$ in (37), the definition of $\mathbb{A}$ in (30) combined with (44) and the following three inequalities (for which the proofs will be given in Section 5.2). Firstly, for the free energy at the final time T, we obtain from Proposition 5.5 and

the continuity of X_T (the evaluation of the path at the final time $t = T$) that

$$\liminf_{L\to\infty} \frac{1}{L^d}\mathcal{F}^V_{L,\alpha}(\mu^L_T) \geq \mathcal{F}^V_\alpha(\rho_T). \tag{50}$$

Secondly,

$$\liminf_{L\to\infty} \frac{1}{L^d}\int_0^T \Psi_L\big(\mu^L_t, \jmath^L_t\big)\,\mathrm{d}t \geq \mathcal{E}\big((\rho_t)_{t\in[0,T]}\big), \tag{51}$$

which follows from Proposition 5.9, and thirdly

$$\liminf_{L\to\infty} \frac{1}{L^d}\int_0^T \Psi^\star_L\big(\mu^L_t, F^V(\mu^L_t)\big)\,\mathrm{d}t \geq \mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big), \tag{52}$$

which is proved in Proposition 5.11. Proposition 4.1 and Proposition 4.4 then yield (48) and (49), respectively. Proposition 4.3 further shows that the path is 2-absolutely continuous in the Wasserstein sense (see (63) in Section 4), from which we can deduce the narrow continuity using Lemma 4.2. The inequality (47) for the free energy at any time $t \in [0,T]$ then follows from another application of Proposition 5.5. □

We now consider a special case for the observed process P_L. We keep the reference process P^V_L as outlined in Section 3.3.1 and consider for some (possibly time-dependent) potential $\tilde{H} \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ the process $P_L = P^{\tilde{V}}_L$ for the potential $\tilde{V}_t = V + \tilde{H}_t$ as defined in Section 2.2.1. Note that both processes have the same initial condition μ^L_0 and their transition rates $r^{V+\tilde{H}_t}$ and r^V coincide up to a change of the external potential (i.e. the functions g_1 and g_2 in (1) coincide for both processes). We assume that the corresponding path measures $(Q^{V+\tilde{H}}_L)_{L\in\mathbb{N}}$ satisfy, as in Section 2.3.2 above, a hydrodynamic limit with hydrodynamic equation

$$\dot{\rho}_t = \Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla(V+\tilde{H}_t)). \tag{53}$$

In this case one can improve the result (46) from Theorem 3.4 by showing that the action functionals $\mathbb{A}^V_L(Q^{V+\tilde{H}}_L)$ converge, as described by the following second main theorem.

Theorem 3.5. *Assume that $P_L = P^{V+\tilde{H}}_L$ for some $\tilde{H} \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ and that $(P_L)_{L\in\mathbb{N}}$ satisfies the assumptions in Theorem 3.4. Moreover, assume that the density of the path $(\pi_t)_{t\in[0,T]}$ is a weak solution to* (53)*, in the sense of* (25)*. Then*

$$\lim_{L\to\infty}\frac{1}{L^d}\mathbb{A}^V_L\big(Q^{V+\tilde{H}}_L\big) = \frac{1}{4}\int_0^T \big\|\nabla\tilde{H}_t\big\|^2_{\chi(\rho_t)}\,\mathrm{d}t = \frac{1}{4}\int_0^T \big\|\dot{\rho}_t - \Delta\phi(\rho_t) - \nabla\cdot(\chi(\rho_t)\nabla V)\big\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t. \tag{54}$$

We postpone the proof of Theorem 3.5 to Section 5.3 below. See also Section 10 in [22] for the specific calculations for the simple exclusion process, which can be seen as a special case of our computations. We further stress that for measures of the form $(P^{\tilde{V}}_L)_{L\in\mathbb{N}}$ the assumption on (45) in Theorem 3.4 is satisfied trivially, since $\mathbb{A}^{\tilde{V}}_L\big(Q^{\tilde{V}}_L\big) = 0$.

Recall that the lower bound (46) in Theorem 3.4 and the limit (54) in Theorem 3.5 coincide (by (40)) if and only if the chain rule (39) holds. The validity of the chain rule (39) for the path $(\pi_t)_{t\in[0,T]}$ in Theorem 3.5 can be shown to be equivalent to the case where the limits in (47), (48) and (49) exist and all three inequalities are equalities.

Theorem 3.6. *Let the assumptions in Theorem 3.5 hold. Further assume that $\mathcal{F}^V_\alpha$ satisfies the chain rule* (39) *for the path $(\rho_t)_{t\in[0,T]}$. Then the free energy converges for all $t \in [0,T]$,*

$$\lim_{L\to\infty}\frac{1}{L^d}\mathcal{F}^V_{L,\alpha}\big(\mu^L_t\big) = \mathcal{F}^V_\alpha(\rho_t). \tag{55}$$

Moreover,

$$\lim_{L\to\infty}\frac{1}{L^d}\int_0^T \Psi_L\big(\mu_t^L, j_t^L\big)\,\mathrm{d}t = \frac{1}{2}\int_0^T \|\dot\rho_t\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t \tag{56}$$

and

$$\lim_{L\to\infty}\frac{1}{L^d}\int_0^T \Psi_L^\star\big(\mu_t^L, F^V(\mu_t^L)\big)\,\mathrm{d}t = \frac{1}{2}\int_0^T \|\Delta\phi(\rho_t)+\nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t. \tag{57}$$

Also the opposite implication holds: If (55), (56) *and* (57) *are satisfied, then* $\mathcal{F}_\alpha^V$ *satisfies the chain rule* (39) *for* $(\rho_t)_{t\in[0,T]}$.

Proof. This proof is similar to calculations performed in [23] and [18], where the authors establish (55) for the hydrodynamic limit of the simple exclusion process. Note that (54), (38), (40) and the chain rule (39) imply

$$\begin{aligned}\lim_{L\to\infty}\frac{1}{L^d}&\bigg(\mathcal{F}_{L,\alpha}^V\big(\mu_T^L\big)+\int_0^T \Psi_L\big(\mu_t^L, j_t^L\big)\,\mathrm{d}t+\int_0^T \Psi_L^\star\big(\mu_t^L, F^V(\mu_t^L)\big)\,\mathrm{d}t\bigg)\\ &= \mathcal{F}_\alpha^V(\rho_T)+\frac{1}{2}\int_0^T \|\dot\rho_t\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t+\frac{1}{2}\int_0^T \|\Delta\phi(\rho_t)+\nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t.\end{aligned}$$

We apply the inequality $\limsup_{n\to\infty}(a_n+b_n+c_n)\geq\limsup_{n\to\infty}a_n+\liminf_{n\to\infty}b_n+\liminf_{n\to\infty}c_n$ to the expression on the left hand side to obtain the inequality

$$\limsup_{L\to\infty}\frac{1}{L^d}\mathcal{F}_{L,\alpha}^V\big(\mu_T^L\big)\leq\mathcal{F}_\alpha^V(\rho_T).$$

The result for an arbitrary time $t\in[0,T]$ then follows for repeating the above proof for the time interval $[0,t]$. The remaining two limits (56) and (57) follow in a similar way by a slight modification of the above steps.

For the opposite implication, we assume that (55), (56) and (57) hold. In this case we have

$$\begin{aligned}\frac{1}{2}\int_0^T &\big\|\dot\rho_t-\Delta\phi(\rho_t)-\nabla\cdot(\chi(\rho_t)\nabla V)\big\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t\\ &= \mathcal{F}_\alpha^V(\rho_T)-\mathcal{F}_\alpha^V(\rho_0)+\frac{1}{2}\int_0^T \|\dot\rho_t\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t+\frac{1}{2}\int_0^T \|\Delta\phi(\rho_t)+\nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t,\end{aligned}$$

which is equivalent to (39) for $t_1=0$ and $t_2=T$. Repeating the above steps for $[0,t]$ (for any $t\in[0,T]$) then finishes the proof. □

Remark on Chain Rule In summary, we have seen that there are at least three ways to verify the chain rule (39). One way is to prove the assumptions of Theorem 3.2. Alternatively, one can derive a Large Deviation Principle, as in Macroscopic Fluctuation Theory (cf. the discussion below Theorem 3.2); or one can directly calculate the limits in Theorem 3.6.

3.4 Examples

Standard examples of particle models described by the class of models in Section 2.1 are (i) the zero-range process (ZRP) for which $\Omega_L=\mathbb{N}_0^{\mathbb{T}_L^d}$, and g_1 is a function that satisfies $g_1(0)=0$ and $g_2=1$; and (ii) the (symmetric) simple exclusion process (SEP), where $\Omega_L=\{0,1\}^{\mathbb{T}_L^d}$, $g_1(n)=\mathbf{1}_{\{n=1\}}$ and $g_2(n)=\mathbf{1}_{\{n=0\}}$; and (iii) the generalised exclusion processes, where $\Omega_L=\{0,\cdots,m\}^{\mathbb{T}_L^d}$, $g_1(n)=\mathbf{1}_{\{n\geq1\}}$ and $g_2(n)=\mathbf{1}_{\{n\leq m\}}$ for some fixed $m\in\mathbb{N}$ [22]. The latter is an example of a non-gradient system. We focus on the two gradient models ZRP and SEP, which have $\mathrm{d}(k)=g_1(k)$ and $\mathrm{d}(k)=k$,

respectively.

3.4.1 Zero-Range Process

The ZRP satisfies the assumptions of Section 2.4.2 if we assume that the rates are strictly monotonically increasing and sub-linear. That is, we assume that there exists $g^* > 0$ such that $0 < g_1(k+1) - g_1(k) \leq g^*$. Since $g_1(0) = 0$ we have $g_1(k) \leq g^* k$. The mobility for the ZRP is given by $\chi(a) = \phi(a)$, where $E_{\nu_\alpha}[g_1(\eta(0))] = \phi(\alpha)$. The reference measure is $\nu_{*,1}(n) = 1/(\prod_{k=1}^{n} g(k))$ and the α-dependent invariant distribution is for $z(\phi(\alpha)) := \sum_{n=0}^{\infty} \phi(\alpha)^n \nu_{*,1}(n)$ given by

$$\nu_{\alpha,1}(\eta(0)) = \frac{\phi(\alpha)^{\eta(0)}}{z(\phi(\alpha))} \nu_{*,1}(\eta(0)).$$

Finally, the free energy is

$$\mathcal{F}_\alpha^V(\rho) = \int_\Lambda \left[\rho(u) \log\left(\frac{\phi(\rho(u))}{\mathrm{e}^{-V(u)}\phi(\alpha)} \right) - \log\left(\frac{z(\phi(\rho(u)))}{z(\mathrm{e}^{-V(u)}\phi(\alpha))} \right) \right] \mathrm{d}u$$

for $f(a) = \rho \log \phi(a) - \log z(\phi(a))$ and $\bar{\rho}_{\alpha,V}(u) = \phi^{-1}(\mathrm{e}^{-V(u)}\phi(\alpha))$.

These considerations establish that Theorems 3.4 to 3.6 can be applied to the ZRP. We now consider the implications of these theorems for hydrodynamic limits. We first compare the path measures for the ZRP (that is, the sequence of P_L^V indexed by L) with some sequence of path measures P_L which concentrate on an absolutely continuous path $(\pi_t)_{t\in[0,T]}$ and satisfies the assumptions of Theorem 3.3. In this case one may apply Theorem 3.4, which establishes an asymptotic lower bound on the rescaled microscopic action $L^{-d}\mathbb{A}_L^V(Q_L)$. If $(\pi_t)_{t\in[0,T]}$ is the hydrodynamic limit of the ZRP then P_L^V has to concentrate on $(\pi_t)_{t\in[0,T]}$, but one also has (in general) that $L^{-d}\mathbb{A}_L^V(Q_L^V) = 0$. Hence, if $L^{-d}\mathbb{A}_L^V(Q_L)$ is bounded away from zero then the path $(\pi_t)_{t\in[0,T]}$ associated to P_L can be ruled out as a possible hydrodynamic limit.

In fact the hydrodynamic limit of the ZRP is known to be given by (52) with $\tilde{H} = 0$ (see Section 5 in [22]), in which case Theorem 3.4 bounds the macroscopic action by zero: $\mathbb{A}((\pi_t)_{t\in[0,T]}) \leq 0$. However this bound is not yet sufficient to show that P_L^V concentrates on $(\pi_t)_{t\in[0,T]}$, so it does not prove the hydrodynamic limit.

We now restrict our consideration to measures of the form $P_L = P_L^{V+\tilde{H}}$ that concentrate on paths which satisfy (53), for some $\tilde{H}$. In this case, Theorem 3.5 may be applied. This establishes that the limit of $L^{-d}\mathbb{A}_L^V(Q_L^{V+\tilde{H}})$ exists. We moreover can verify the assumptions of Theorem 3.2 (at least for $d = 1$) or alternatively rely on the existence of the pathwise LDP (see [3]), which shows that also Theorem 3.6 holds – this establishes a lower bound $\mathbb{A}((\pi_t)_{t\in[0,T]}) \geq 0$ for any path $(\pi_t)_{t\in[0,T]}$ that solves (53), with some $\tilde{H}$. This means that $(\pi_t)_{t\in[0,T]}$ is only admissible as a candidate for the hydrodynamic limit of the ZRP, if it is a (weak) solution to (53) with $\tilde{H} = 0$ (otherwise one has the contradiction $0 = \lim_{L\to\infty} L^{-d}\mathbb{A}_L^V(Q_L^V) = \mathbb{A}((\pi_t)_{t\in[0,T]}) > 0$).

3.4.2 Simple Exclusion Process

For the SEP the invariant reference measure is $\nu_{*,1}(0) = \nu_{*,1}(1) = 1$ and the α-dependent invariant product measure are Bernoulli distributed $\nu_{\alpha,1}(\eta(0)) = \alpha^{\eta(0)}(1-\alpha)^{1-\eta(0)}$. The functions ϕ and χ are given by $\phi(\alpha) = \alpha$ and $\chi(\alpha) = \alpha(1-\alpha)$. The free energy is given by

$$\mathcal{F}_\alpha^V(\rho) = \int_\Lambda \left[\rho(u) \log\left(\frac{\rho(u)}{\alpha \mathrm{e}^{-V(u)}} \right) + (1-\rho(u)) \log\left(\frac{1-\rho(u)}{1-\alpha} \right) + \log\Big(\alpha \mathrm{e}^{-V(u)} + (1-\alpha) \Big) \right] \mathrm{d}u,$$

which is of the form (22) for the free energy density $f(a) = a\log a + (1-a)\log(1-a)$ and the stationary density is $\bar{\rho}_{\alpha,V}(u) = \alpha \mathrm{e}^{-V(u)}/(\alpha \mathrm{e}^{-V(u)} + (1-\alpha))$.

For the sequence $P_L^{V+\tilde{H}}$ the hydrodynamic limit is again given in (53), which has for suitable initial condition a unique weak solution (see Proposition 5.1 on page 273 in [22]). We can proceed as for the ZRP and can establish (under suitable assumptions) that the results of Theorem 3.4 and Theorem 3.5 hold.

Note that this process does not satisfy the assumptions of Theorem 3.2 (as the assumption $\chi'(a) \geq C_*$ is not satisfied). Nonetheless, we can establish the chain rule (39) if the pathwise LDP holds (cf. the discussion at the end of Section 3.1). This was e.g. proved in [22, Chapter 10] (see also [6]), such that also in this case the results of Theorem 3.6 hold.

4 Regularity of Paths and the Chain Rule

The main aim of this section is to prove Theorem 3.2. The central difficulty is that classical approaches to establish chain rules in metric spaces rely on λ-convexity of the functional under consideration; this property is delicate and apparently not sufficiently well understood in a context other than the classic (unweighted) Wasserstein setting. The process considered here are, however, naturally linked to weighted Wasserstein spaces, where important elements of the classic Wasserstein theory are still missing. We circumvent this problem by showing that while the classic Wasserstein space is not the natural space for the processes we study, they can be cast in this setting. The analysis is then somewhat technical, but follows largely arguments in [1]. The novel Ψ-$\Psi^\star$-structure is thus less relevant in this section than for the proofs in Section 5.

In the following, we consider paths with conserved volume, for which also the action is finite: $\mathbb{A}((\rho_t)_{t\in[0,T]}) < \infty$. Combined with $\mathcal{F}_\alpha^V(\rho_0) < \infty$ and (30), this implies that $\mathcal{E}((\rho_t)_{t\in[0,T]}) < \infty$ and $\mathcal{E}^\star((\rho_t)_{t\in[0,T]}) < \infty$. We will see that the former of the two implies regularity in time (that $(\rho_t)_{t\in[0,T]}$ is absolutely-continuous in the Wasserstein sense) and the latter yields certain regularity in space (such that e.g. the weak gradient $\nabla\phi(\rho)$ exists a.e. in Λ).

The following steps are based on ideas from Section 4 in [10]. For a more recent and concise representation of the following material, we refer to Appendices D.5 and D.6) in [19]. A discussion of similar content in terms of interacting particle systems can e.g. be found in [6].

For any topological space $\mathcal{S}$, we denote with $\mathscr{D}(\mathcal{S};\mathbb{R}) = C_c^\infty(\mathcal{S};\mathbb{R})$ the vector space of real-valued infinitely often differentiable and compactly supported functions on $\mathcal{S}$ and equip $\mathscr{D}(\mathcal{S};\mathbb{R})$ with the usual topology for test functions, see e.g. [19, Appendix D.1]. Its topological dual, the space of (Schwartz) distributions, will be denoted with $\mathscr{D}'(\mathcal{S};\mathbb{R})$. The application of $g \in \mathscr{D}(\mathcal{S};\mathbb{R})$ to a distribution $\vartheta \in \mathscr{D}'(\mathcal{S};\mathbb{R})$ is denoted by $\langle \vartheta, g\rangle$.

The Otto calculus yields a formal interpretation of $\mathcal{M}_+(\Lambda)$ as an infinite dimensional Riemannian manifold (see for example Chapter 15 in [39] or Section 8.1.2 in [38]). For a measure $\pi \in \mathcal{M}_+(\Lambda)$, one can define three isometric spaces $H_\pi^1(\Lambda;\mathbb{R})$, $H_\pi^{-1}(\Lambda;\mathbb{R})$ and $\mathcal{L}^2_{\nabla,\pi}(\Lambda;\mathbb{R}^d)$, which all can play the role of the 'tangent space' at π. We next give precise definitions of all three spaces. For $h\colon \Lambda \to \mathbb{R}^d$, we define the norm $\|h\|_\pi^2 := \int_\Lambda |h(u)|^2 \pi(\mathrm{d}u)$. For $g \in W^1_{\mathrm{loc}}(\Lambda;\mathbb{R})$ this norm gives rise to the semi-norm $\|g\|_{1,\pi} := \|\nabla g\|_\pi$, where ∇g denotes the weak derivative of g. Since $\{g \in \mathscr{D}(\Lambda;\mathbb{R}) : \int_\Lambda g\,\mathrm{d}u = 0\}$ equipped with $\|\cdot\|_{1,\pi}$ is a normed space, we can define its completion to be $H_\pi^1(\Lambda;\mathbb{R})$. For $\vartheta \in \mathscr{D}'(\Lambda;\mathbb{R})$ the dual norm, which is defined as

$$\|\vartheta\|_{-1,\pi}^2 := \sup_{g\in H_\pi^1(\Lambda;\mathbb{R})} \big(2\langle \vartheta, g\rangle - \|g\|_{1,\pi}^2\big), \tag{58}$$

gives rise to $H_\pi^{-1}(\Lambda;\mathbb{R}) := \{\vartheta \in \mathscr{D}'(\Lambda;\mathbb{R}) : \|\vartheta\|_{-1,\pi} < \infty\}$, the dual of $H_\pi^1(\Lambda;\mathbb{R})$. Note that $H_\pi^1(\Lambda;\mathbb{R})$

is a Hilbert space (with inner product $\langle\cdot,\cdot\rangle_{1,\pi}$ defined in the obvious way using the polarisation identity for inner products); it therefore is reflexive, which implies the existence of a linear and isometric map from $H^1_\pi(\Lambda;\mathbb{R})$ to $H^{-1}_\pi(\Lambda;\mathbb{R})$, formally given by $g \mapsto -\nabla\cdot(\pi\nabla g)$. The inner product on $H^{-1}_\pi(\Lambda;\mathbb{R})$ will be denoted with $\langle\cdot,\cdot\rangle_{-1,\pi}$. Finally, let $\mathcal{L}^2_{\nabla,\pi}(\Lambda;\mathbb{R}^d)$ be the completion of $\{\nabla\zeta : \zeta\in\mathscr{D}(\Lambda;\mathbb{R})\}$ with respect to $\|\cdot\|_\pi$. It is then easy to see that $H^1_\pi(\Lambda;\mathbb{R})$ is also isometric to $\mathcal{L}^2_{\nabla,\pi}(\Lambda;\mathbb{R}^d)$ (cf. page 379 in [19]). We will denote the map from $H^1_\pi(\Lambda;\mathbb{R})$ to $\mathcal{L}^2_{\nabla,\pi}(\Lambda;\mathbb{R}^d)$ with ∇.

For our purposes, the spaces $H^{-1}_\pi(\Lambda;\mathbb{R})$ and $\mathcal{L}^2_{\nabla,\pi}(\Lambda;\mathbb{R}^d)$ yield the more relevant representations. The two prominent cases that will appear in the following are $\pi(\mathrm{d}u)=\rho(u)\mathrm{d}u$ and $\pi(\mathrm{d}u)=\chi(\rho(u))\mathrm{d}u$. In these cases we will identify the densities ρ and $\chi(\rho)$ as measures and write $H^1_\rho(\Lambda;\mathbb{R})$ and $H^1_{\chi(\rho)}(\Lambda;\mathbb{R})$ instead of $H^1_\pi(\Lambda;\mathbb{R})$ (and similar for the other spaces we just introduced).

4.1 Regularity of Paths on the Hydrodynamic Scale

Now, fix a path $(\pi_t)_{t\in[0,T]}\in\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$ that is absolutely continuous with respect to the Lebesgue measure with density $(\rho_t)_{t\in[0,T]}$. We equip $C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ with the $(\rho_t)_{t\in[0,T]}$ dependent semi-norm $G\mapsto(\int_0^T\|\nabla G_t\|^2_{\chi(\rho_t)}\mathrm{d}t)^{1/2}$, on which we define the two real valued linear operators

$$L_{\mathcal{E}}(G) := \int_\Lambda \rho_T G_T\,\mathrm{d}u - \int_\Lambda \rho_0 G_0\,\mathrm{d}u - \int_0^T\int_\Lambda \rho_t\,\partial_t G_t\,\mathrm{d}u\,\mathrm{d}t$$

and

$$L_{\mathcal{E}^\star}(G) := \int_0^T\int_\Lambda \phi(\rho_t)\nabla\cdot\nabla G_t\,\mathrm{d}u\,\mathrm{d}t - \int_0^T\int_\Lambda \chi(\rho_t)\nabla V\cdot\nabla G_t\,\mathrm{d}u\,\mathrm{d}t.$$

Note that these two operators coincide with the left and right hand side of (25), respectively. Moreover, the corresponding operator norms are given by $\mathcal{E}((\rho_t)_{t\in[0,T]})$ in (28) and $\mathcal{E}^\star((\rho_t)_{t\in[0,T]})$ in (29), respectively (cf. e.g. [10, 19]).

Under the assumptions of Theorem 3.4, we have prior information on the regularity of the path $(\rho_t)_{t\in[0,T]}$, i.e. we can assume that $\mathcal{E}((\rho_t)_{t\in[0,T]}),\mathcal{E}^\star((\rho_t)_{t\in[0,T]})<\infty$ (such that $L_{\mathcal{E}}$ and $L_{\mathcal{E}^\star}$ are bounded linear operators).

Note that $L_{\mathcal{E}}$ and $L_{\mathcal{E}^\star}$ are both invariant under addition of a constant in the sense that $L_{\mathcal{E}^\star}(G)=L_{\mathcal{E}^\star}(G+c)$ for any $c\in\mathbb{R}$. We thus can (with slight abuse of notation) redefine $L_{\mathcal{E}}$ and $L_{\mathcal{E}^\star}$ as operators on $\{\nabla G : G\in C^{1,2}([0,T]\times\Lambda;\mathbb{R})\}$, equipped with $\nabla G\mapsto(\int_0^T\|\nabla G_t\|^2_{\chi(\rho_t)}\mathrm{d}t)^{1/2}$, as

$$L_{\mathcal{E}}(\nabla G) := L_{\mathcal{E}}(G) \quad\text{and}\quad L_{\mathcal{E}^\star}(\nabla G) := L_{\mathcal{E}^\star}(G).$$

Let $\mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$ be the $(\rho_t)_{t\in[0,T]}$ dependent completion of $\{\nabla G : G\in C^{1,2}([0,T]\times\Lambda;\mathbb{R})\}$ with respect to $\nabla G\mapsto(\int_0^T\|\nabla G_t\|^2_{\chi(\rho_t)}\mathrm{d}t)^{1/2}$. Note that if $h=(h_t)_{t\in[0,T]}\in\mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$, then $h_t\in\mathcal{L}^2_{\nabla,\chi(\rho_t)}(\Lambda;\mathbb{R}^d)$ for a.a. $t\in[0,T]$. In Section 4.2 we will also consider $\mathcal{L}^2_{\nabla,\mathrm{id}}([0,T]\times\Lambda;\mathbb{R}^d)$, where the norm is replaced with $\nabla G\mapsto(\int_0^T\|\nabla G_t\|^2_{\rho_t}\mathrm{d}t)^{1/2}$.

Since $\mathcal{E}((\rho_t)_{t\in[0,T]}),\mathcal{E}^\star((\rho_t)_{t\in[0,T]})<\infty$ the Bounded Linear Transformation Theorem (see e.g. Theorem I.6 in [35]), allows us to extend $L_{\mathcal{E}}(\nabla G)$ and $L_{\mathcal{E}^\star}(\nabla G)$ to bounded linear operators on $\mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$ with the same operator norms as above. For $h\in\mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$ we have

$$L_{\mathcal{E}}(h) = \int_\Lambda \rho_T\nabla^{-1}h_T\,\mathrm{d}u - \int_\Lambda \rho_0\nabla^{-1}h_0\,\mathrm{d}u - \int_0^T\int_\Lambda \rho_t\,\partial_t(\nabla^{-1}h_t)\,\mathrm{d}u\,\mathrm{d}t,$$

where ∇^{-1} denotes (for each $t\in[0,T]$) the isometric map from $\mathcal{L}^2_{\nabla,\chi(\rho_t)}(\Lambda;\mathbb{R}^d)$ to $H^1_{\chi(\rho_t)}(\Lambda;\mathbb{R})$.

Further

$$L_{\mathcal{E}^\star}(h) = \int_0^T \int_\Lambda \phi(\rho_t)\nabla\cdot h_t \,\mathrm{d}u\,\mathrm{d}t - \int_0^T \int_\Lambda \chi(\rho_t)\nabla V\cdot h_t \,\mathrm{d}u\,\mathrm{d}t.$$

By Riesz' representation theorem (e.g. Theorem II.4 in [35]), there exist unique elements $v, w \in \mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$, with $v = (v_t)_{t\in[0,T]}$ and $w = (w_t)_{t\in[0,T]}$, for which these two bounded operators can be represented by

$$L_{\mathcal{E}}(h) = \int_0^T \int_\Lambda \chi(\rho_t)v_t\cdot h_t \,\mathrm{d}u\,\mathrm{d}t, \qquad L_{\mathcal{E}^\star}(h) = \int_0^T \int_\Lambda \chi(\rho_t)w_t\cdot h_t \,\mathrm{d}u\,\mathrm{d}t. \tag{59}$$

Substituting (59) in (28) and (29) yields (c.f. Lemma 4.8 in [10])

$$\mathcal{E}\big((\rho_t)_{t\in[0,T]}\big) = \frac{1}{2}\int_0^T \|v_t\|^2_{\chi(\rho_t)}\,\mathrm{d}t, \qquad \mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big) = \frac{1}{2}\int_0^T \|w_t\|^2_{\chi(\rho_t)}\,\mathrm{d}t. \tag{60}$$

Proposition 4.1. *Assume that $\mathcal{E}((\rho_t)_{t\in[0,T]}) < \infty$ and that χ satisfies the assumptions of Section 2.4.2. Then the weak time derivative of ρ_t, denoted $\dot\rho_t$, exists in $H^{-1}_{\chi(\rho_t)}(\Lambda;\mathbb{R})$ for a.a. $t\in[0,T]$. Moreover,*

$$\mathcal{E}\big((\rho_t)_{t\in[0,T]}\big) = \frac{1}{2}\int_0^T \|\dot\rho_t\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t. \tag{61}$$

Proof. Results of this kind are standard and we hence only sketch the proof. Consider the unique $v\in\mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$ from (59) and recall that $v_t\in\mathcal{L}^2_{\nabla,\chi(\rho_t)}(\Lambda;\mathbb{R}^d)$ for a.a. $t\in[0,T]$.

Following e.g. Lemma 4.8 in [10] (see also [13]), one shows that $\mathcal{E}((\rho_t)_{t\in[0,T]})<\infty$ implies that $t\mapsto\langle\rho_t,\cdot\rangle$ is absolutely continuous in the sense of distributions, such that the distributional derivative $\dot\rho_t\in\mathscr{D}'(\Lambda;\mathbb{R})$ exists for a.a. $t\in(0,T)$. In our case, the latter satisfies for $G\in\mathscr{D}(\Lambda;\mathbb{R})$ and a.a. $t\in(0,T)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Lambda \rho_t G\,\mathrm{d}u = \langle\dot\rho_t, G\rangle = \int_\Lambda \chi(\rho_t)v_t\cdot\nabla G\,\mathrm{d}u. \tag{62}$$

Thus $\dot\rho_t = -\nabla\cdot(\chi(\rho_t)v_t)$ in the distributional sense for a.a. $t\in(0,T)$, such that $v_t\in\mathcal{L}^2_{\nabla,\chi(\rho_t)}(\Lambda;\mathbb{R}^d)$ can uniquely be identified with $\dot\rho_t$. Further the isometry from $\mathcal{L}^2_{\nabla,\chi(\rho_t)}(\Lambda;\mathbb{R}^d)$ to $H^{-1}_{\chi(\rho_t)}(\Lambda;\mathbb{R})$ (for a.a. $t\in[0,T]$) implies that $\dot\rho_t\in H^{-1}_{\chi(\rho_t)}(\Lambda;\mathbb{R})$ and (61) also follows. □

Let $p\in[1,\infty]$. We say a path $(\pi_t)_{t\in[0,T]}$ is *p-absolutely continuous (in the Wasserstein sense)*, if there exists a function $m\in\mathcal{L}^p([0,T];\mathbb{R})$, such that for any $0\le t_1<t_2\le T$

$$W_2(\pi_{t_1},\pi_{t_2}) \le \int_{t_1}^{t_2} m(s)\,\mathrm{d}s, \tag{63}$$

where W_2 denotes the 2-Wasserstein distance [38, 1]. In this case, the metric derivative (cf. equation (1.1.3) in [1]) exists for a.a. $t\in(0,T)$,

$$|\pi_t'| := \limsup_{h\to 0}\left(\frac{W_2(\pi_t,\pi_{t+h})}{h}\right) < \infty$$

and $t\mapsto|\pi_t'|$ is the minimal function that satisfies (63), see Theorem 1.1.2 in [1]. In other words, $(\pi_t)_{t\in[0,T]}$ is p-absolutely continuous if and only if the map $t\mapsto|\pi_t'|$ is an element of $\mathcal{L}^p([0,T];\mathbb{R})$. From now on we consider the case $p=2$.

Lemma 4.2. *A path $(\pi_t)_{t\in[0,T]}\in\mathcal{D}([0,T];\mathcal{M}_+(\Lambda))$ is 2-absolutely continuous if and only if there exists a vector field $\tilde v=(\tilde v_t)_{t\in[0,T]}$ with $\tilde v_t\in\mathcal{L}^2_{\nabla,\pi_t}(\Lambda;\mathbb{R}^d)$ and $\int_0^T\|\tilde v_t\|_{\pi_t}\,\mathrm{d}t<\infty$ that satisfies $\dot\pi_t + \nabla\cdot(\pi_t\tilde v_t)=0$ in the distributional sense for almost all $t\in[0,T]$. In this case we have in particular $(\pi_t)_{t\in[0,T]}\in C([0,T];\mathcal{M}_+(\Lambda))$.*

Proof. The result follows from a modification of Lemma 8.1.2 and Theorem 8.3.1 in [1] to the domain Λ. Assume first that $(\pi_t)_{t\in[0,T]}$ is 2-absolutely continuous. Then Theorem 8.3.1 implies that the continuity equation $\dot\pi_t + \nabla\cdot(\pi_t\tilde v_t) = 0$ holds for some $\tilde v_t$, which can, by Lemma 8.4.2 in [1], without loss of generality be chosen to satisfy $\tilde v_t \in \mathcal{L}^2_{\nabla,\pi_t}(\Lambda;\mathbb{R}^d)$.

For the opposite implication we assume that the continuity equation holds and that moreover $\int_0^T \|\tilde v_t\|_{\pi_t}\,\mathrm{d}t < \infty$. An application of the Hölder inequality combined with $\sup_{t\in[0,T]}\pi_t(\Lambda) < \infty$ ensures that $\int_0^T\int_\Lambda |\tilde v_t(u)|\,\pi_t(\mathrm{d}u)\,\mathrm{d}t < \infty$. Lemma 8.1.2 thus implies that the curve has a weakly continuous modification $(\tilde\pi_t)_{t\in[0,T]} \in C([0,T];\mathcal{M}_+(\Lambda))$. Now, since every right-continuous path that admits a continuous modification already has to be continuous, we have $(\pi_t)_{t\in[0,T]} = (\tilde\pi_t)_{t\in[0,T]}$. This allows us to apply the reverse implication of Theorem 8.3.1 to $(\pi_t)_{t\in[0,T]}$, which yields that $(\pi_t)_{t\in[0,T]}$ is 2-absolutely continuous. □

The Wasserstein distance W_2 has a fluid dynamical representation in terms of the Brenier-Benamou formula (compare Equation (8.0.3) in [1] and Section 8.1 in [38]). The distance of two measures $\pi,\hat\pi\in\mathcal{M}_+(\Lambda)$ with $\pi(\Lambda)=\hat\pi(\Lambda)>0$ is given by

$$W_2^2(\pi,\hat\pi) = \inf\Big\{\int_0^1 \|\tilde v_t\|^2_{\mu_t}\,\mathrm{d}t \;\Big|\; \mu_0=\pi,\ \mu_1=\hat\pi,\ \dot\mu_t+\nabla\cdot(\mu_t\tilde v_t)=0\Big\},$$

where the infimum is taken over all 2-absolutely continuous paths of measures $(\mu_t)_{t\in[0,T]}$ and velocities $\tilde v_t\in\mathcal{L}^2_{\nabla,\mu_t}(\Lambda;\mathbb{R}^d)$ satisfying the continuity equation above.

Let $(\pi_t)_{t\in[0,T]}$ be absolutely continuous with respect to the Lebesgue measure with density $(\rho_t)_{t\in[0,T]}$. We say that $(\rho_t)_{t\in[0,T]}$ is 2-absolutely continuous if $(\pi_t)_{t\in[0,T]}$ is 2-absolutely continuous. Moreover, we will identify densities with their associated measures. In particular, we write $W_2^2(\rho,\hat\rho) = W_2^2(\pi,\hat\pi)$ for $\pi(\mathrm{d}u)=\rho(u)\mathrm{d}u$ and $\pi(\mathrm{d}u)=\rho(u)\mathrm{d}u$.

Proposition 4.3. *Assume that $\mathcal{E}((\rho_t)_{t\in[0,T]}) < \infty$ and that χ satisfies the assumptions of Section 2.4.2. Then $(\rho_t)_{t\in[0,T]}$ is 2-absolutely continuous in the Wasserstein sense.*

Proof. We choose the time rescaling $\bar t = t(t_2-t_1)+t_1$ and set $\mu_t=\rho_{\bar t}$ and $\tilde v_t = (t_2-t_1)(\chi(\rho_{\bar t})v_{\bar t})/\rho_{\bar t}$, such that $\dot\mu_t+\nabla\cdot(\mu_t\tilde v_t)=0$ by construction. We obtain for all $0\le t_1<t_2\le T$

$$W_2^2(\rho_{t_1},\rho_{t_2}) \le (t_2-t_1)\int_{t_1}^{t_2}\|(\chi(\rho_t)v_t)/\rho_t\|^2_{\rho_t}\,\mathrm{d}t \le (t_2-t_1)\int_{t_1}^{t_2} C_{\mathrm{Lip}}\|v_t\|^2_{\chi(\rho_t)}\,\mathrm{d}t<\infty,$$

such that the metric derivative satisfies for almost all $t\in[0,T)$

$$|\rho_t'| = \limsup_{h\to0}\left(\frac{W_2(\rho_t,\rho_{t+h})}{h}\right) \le \sqrt{C_{\mathrm{Lip}}}\|v_t\|_{\chi(\rho_t)}. \tag{64}$$

The square integrability of the right hand side now implies that $(\rho_t)_{t\in[0,T]}$ is 2-absolutely continuous. □

Proposition 4.4. *Assume that $\mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big) < \infty$ and that f,ϕ and χ satisfy the assumptions of Section 2.4.2. Then*

$$\mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big) = \frac12\int_0^T \|\Delta\phi(\rho_t)+\nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t = \frac12\int_0^T\|f''(\rho_t)\nabla\rho_t+\nabla V\|^2_{\chi(\rho_t)}\,\mathrm{d}t. \tag{65}$$

Proof. $\mathcal{E}^\star\big((\rho_t)_{t\in[0,T]}\big)<\infty$ implies that the distributional derivative of $\phi(\rho_t)\in\mathcal{L}^1_{\mathrm{loc}}(\Lambda;\mathbb{R})$ satisfies $\nabla\phi(\rho_t)\in\mathcal{L}^1_{\mathrm{loc}}(\Lambda;\mathbb{R}^d)$ for a.a. $t\in[0,T]$ (cf. Appendix D.6 in [19]). Equivalently, $\phi(\rho_t)\in W^{1,1}_{\mathrm{loc}}(\Lambda;\mathbb{R})$ for a.a. $t\in[0,T]$. The first identity in (65) can be established as in Appendix D.6 in [19] (for the choice

$\mu(\mathrm{d}u) = \chi(\rho_t(u))\mathrm{d}u$). We turn to the second identity. Since ϕ^{-1} is continuously differentiable with bounded derivative, we obtain by the chain rule for functions in $W^{1,1}_{\mathrm{loc}}(\Lambda;\mathbb{R})$ with bounded derivative (see e.g. Theorem 4 (ii) in [17]) that also $\nabla\rho_t \in \mathcal{L}^1_{\mathrm{loc}}(\Lambda;\mathbb{R})$, and thus $\rho_t \in W^{1,1}_{\mathrm{loc}}(\Lambda;\mathbb{R})$, for almost all $t \in [0,T]$. The derivative is for almost all $u \in \Lambda$ given by

$$\nabla\rho_t(u) = (\phi^{-1})'(\phi(\rho_t(u)))\nabla\phi(\rho_t(u)) = \frac{\nabla\phi(\rho_t(u))}{\phi'(\rho_t(u))}, \tag{66}$$

where the last identity follows from the Implicit Function Theorem. Multiplying with $\phi'(\rho_t)$ and using the local Einstein relation (27) we obtain that almost everywhere

$$\nabla\phi(\rho_t) = \phi'(\rho_t)\nabla\rho_t = \chi(\rho_t)f''(\rho_t)\nabla\rho_t. \tag{67}$$

Combined with w in (60), we have for any $G \in \mathscr{D}(\Lambda;\mathbb{R})$ and almost all $t \in [0,T]$ that

$$\int_\Lambda \chi(\rho_t)w_t \cdot \nabla G\,\mathrm{d}u = \int_\Lambda \big(\nabla\phi(\rho_t) + \chi(\rho_t)\nabla V\big)\cdot\nabla G\,\mathrm{d}u = \int_\Lambda \chi(\rho_t)[f''(\rho_t)\nabla\rho_t + \nabla V]\cdot\nabla G\,\mathrm{d}u$$

such that we can identify $w_t = f''(\rho_t)\nabla\rho_t + \nabla V$. Substituting this identity in (60) yields the final result. □

4.2 Chain Rule for the Free Energy

In this section, we prove Theorem 3.2, which establishes rigorously the validity of the macroscopic chain rule (39), for which we so far gave only a formal derivation. Consider a given path $(\rho_t)_{t\in[0,T]}$ that satisfies $\mathbb{A}((\rho_t)_{t\in[0,T]}) < \infty$. We restrict ourselves to densities $\rho, \hat\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ s.t. $\int_\Lambda \rho\,\mathrm{d}u = \int_\Lambda \hat\rho\,\mathrm{d}u > 0$ and continue to identify densities with measures. The constant volume implies that free energy differences do not depend on α. Indeed, defining $\mathcal{F}(\rho) := \int_\Lambda f(\rho(u))\mathrm{d}u$ and $\mathcal{V}(\rho) := \int_\Lambda V(u)\rho(u)\mathrm{d}u$ (for $V \in C^2(\Lambda;\mathbb{R})$), we can define an α-independent modification of the free energy

$$\mathcal{F}^V(\rho) := \mathcal{F}(\rho) + \mathcal{V}(\rho), \tag{68}$$

which is (with (22)) easily seen to satisfy $\mathcal{F}^V_\alpha(\hat\rho) - \mathcal{F}^V_\alpha(\rho) = \mathcal{F}^V(\hat\rho) - \mathcal{F}^V(\rho)$.

We assume that $f \in C^2([0,\infty);\mathbb{R})$ satisfies the assumptions in Section 2.4.2, such that the functional $\mathcal{F}\colon \mathcal{L}^1(\Lambda;[0,\infty)) \to (-\infty,\infty]$ is proper and lower-semicontinuous (see Remark 9.3.8 in [1]). Note that for $N_{\max} = \infty$ the assumption $\lim_{r\to N_{\max}} f'(r) = \infty$ implies super linearity of f.

We set

$$L_f(a) := af'(a) - f(a) = \int_0^a rf''(r)\,\mathrm{d}r$$

and note the similarity to $\phi(a) = \int_0^a \phi'(r)\,\mathrm{d}r = \int_0^a \chi(r)f''(r)\,\mathrm{d}r$ (where we again used the local Einstein relation (27)); in particular $L_f'(a)/a = f''(a) = \phi'(a)/\chi(a)$. The quantity L_f is sometimes referred to as a 'pressure' function due to its relation to the thermodynamic pressure in classical thermodynamics, see e.g. Remark 5.18 (ii) in [38].

We denote the (2-)Wasserstein distance between ρ and $\hat\rho$ with $W_2(\rho,\hat\rho)$. A constant speed geodesic (connecting ρ to $\hat\rho$) is a curve $(\rho_t)_{t\in[0,1]}$ such that ($\rho_0 = \rho$, $\rho_1 = \hat\rho$ and) $W_2(\rho_s,\rho_t) = |t-s|W_2(\rho,\hat\rho)$ for all $s,t \in [0,T]$. With this, a functional $\mathcal{G}$ is called λ-convex (also called semi-convex) for $\lambda \in \mathbb{R}$ if the inequality

$$\mathcal{G}(\rho_t) \le (1-t)\mathcal{G}(\rho_0) + t\mathcal{G}(\rho_1) - \frac{\lambda}{2}t(1-t)W_2^2(\rho_0,\rho_1) \tag{69}$$

holds for each constant speed geodesic $(\rho_t)_{t\in[0,1]}$. Note that if two functionals $\mathcal{G}_i$ are λ_i-convex for

$i = 1, 2$, then clearly $\mathcal{G}_1 + \mathcal{G}_2$ is λ-convex with $\lambda = \min(\lambda_1, \lambda_2)$.

We call $\mathcal{G}$ geodesically convex if the map $t \mapsto \mathcal{G}(\rho_t)$ is convex for any geodesic $(\rho_t)_{t\in[0,1]}$ (which is equivalent to λ-convexity for $\lambda = 0$). A useful criterion for geodesic convexity of the free energy $\mathcal{F}$ is the McCann condition (see Proposition 9.3.9 and equation (9.3.11) in [1]): A convex function $f \in C^2([0,\infty);\mathbb{R})$ with $f(0) = 0$ satisfies the McCann condition (in d dimensions) if the map

$$s \mapsto s^d f(s^{-d}) \tag{70}$$

is convex on $(0,\infty)$ (cf. the discussion in Section 9.3 in [1]). In the case $d = 1$, convexity of f is sufficient to establish geodesic convexity. For a potential energy of the form $\mathcal{V}(\rho) = \int_\Lambda V(u)\rho(u)\mathrm{d}u$ λ-convexity is equivalent to λ-convexity (also called strong convexity) of V on Λ (see equation (9.3.3) and Proposition 9.3.2 in [1]), which is $V((1-t)x+ty) \le (1-t)V(x)+tV(y)-(\lambda/2)t(1-t)\|x-y\|^2$. For $V \in C^2(\Lambda;\mathbb{R})$ the Hessian matrix is bounded and this assumption is trivially satisfied. Note that under the assumption that $\mathcal{F}$ is geodesically-convex and $\mathcal{V}$ is λ-convex for some $\lambda \le 0$, also $\mathcal{F}^V$ is λ-convex.

4.2.1 Assumptions for Chain Rule

To our knowledge, minimal sufficient conditions for the validity of a chain rule of the form (39) are still an open question. One difficulty is that the existing theory requires λ-convexity of the functional in question. In the case of independent particles (with $\chi(a) = \phi(a) = a$) sufficient conditions for λ-convex functionals can be obtained from the general theory for gradient flows in Wasserstein spaces, which was established in [1] (see also [38, 36]). We note that generalisations of the gradient flow theory in Wasserstein spaces with non-linear (usually concave) mobilities have been considered in the literature, see e.g. [25, 26, 27, 12, 11]. Yet, establishing the chain rule in a weighted Wasserstein metric is fraught with technical difficulties, in particular λ-convexity of the functional. We overcome this difficulty here by showing that in the setting studied here, where a weighted Wasserstein metric is the natural space, the chain rule can be established in an unweighted (classical) Wasserstein setting, where strong tools are available.

In this section, we establish the chain rule (39) in the special case that the density f of the free energy $\mathcal{F}^V$ satisfies the McCann condition for geodesic convexity (70) and the particle process is 'not too far away' from the process with independent particles (where $\chi(a) = \phi(a) = a$): We consider the case $N_{\max} = \infty$ and assume there exists $C_* > 0$ (without loss of generality the same constant which bounds $\phi'(a)$ from below) such that

$$C_* \le \chi'(a) \tag{71}$$

for almost all $a \in (0,\infty)$. This implies that $C_* \le \chi'(a), \phi'(a) \le C_{\mathrm{Lip}}$, such that also $C_* a \le \chi(a), \phi(a) \le C_{\mathrm{Lip}} a$. We obtain for any $\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ that the norms $\|\cdot\|_\rho$ and $\|\cdot\|_{\chi(\rho)}$ are equivalent,

$$C_*\|\cdot\|_\rho \le \|\cdot\|_{\chi(\rho)} \le C_{\mathrm{Lip}}\|\cdot\|_\rho. \tag{72}$$

In this case also the limit points coincide such that $\mathcal{L}^2_{\nabla,\chi(\rho)}(\Lambda;\mathbb{R}^d) = \mathcal{L}^2_{\nabla,\rho}(\Lambda;\mathbb{R}^d)$. This will allow us to leverage results from the classical Wasserstein framework in [1].

Remark. The Lipschitz continuity of $\chi(a)$ implies that $\mathcal{L}^2_{\nabla,\rho}(\Lambda;\mathbb{R}^d) \subseteq \mathcal{L}^2_{\nabla,\chi(\rho)}(\Lambda;\mathbb{R}^d)$. In general, this is a strict inclusion (consider e.g. the case of the SEP with $\chi(a) = a(1-a)$ and $\rho = 1$ on a subset $O \subseteq \Lambda$ with positive Lebesgue measure). A (weaker, density ρ dependent) condition for the opposite inclusion to hold is

$$\inf_{u\in\Lambda} \frac{\chi(\rho(u))}{\rho(u)} > 0,$$

which can in this case replace the constant in the lower bound of (72). Note that this is a density specific condition, whereas the above condition (71) is a model specific condition (which is independent of ρ). For the SEP, this condition is satisfied precisely in the case when ρ is bounded away from the maximal possible local particle density, i.e. $\rho \leq N_{\max} - \epsilon$ (for some $\epsilon > 0$). The same considerations show that in general $\mathcal{L}^2_{\nabla,\mathrm{id}}([0,T]\times\Lambda;\mathbb{R}^d) \subseteq \mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$ and that (71), or alternatively

$$\inf_{(t,u)\in[0,T]\times\Lambda} \frac{\chi(\rho_t(u))}{\rho_t(u)} > 0,$$

ensures that $\mathcal{L}^2_{\nabla,\mathrm{id}}([0,T]\times\Lambda;\mathbb{R}^d) = \mathcal{L}^2_{\nabla,\chi}([0,T]\times\Lambda;\mathbb{R}^d)$.

4.2.2 Validity of the Chain Rule

The following results, which are mainly based on Chapter 9 and 10 in [1], relate $L_f(\rho)$ to the directional derivative, the Fréchet-subdifferential, and the metric slope of $\mathcal{F}(\rho)$. Below we sketch results which can be obtained by a suitable modification of the results in [1]. More precisely, we are interested in the case where the domain is $\Lambda = \mathbb{T}^d$ and the measures of interest are absolutely continuous with respect to the Lebesgue measure.

As shown in Theorem 1.25 in [36] there exists for any $\rho, \hat\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ with $\int_\Lambda \rho\,\mathrm{d}u = \int_\Lambda \hat\rho\,\mathrm{d}u > 0$ a unique optimal transport map from ρ to $\hat\rho$ of the form $r = i - \nabla\varphi$, where φ is semi-concave (i.e. there exists a constant $C > 0$ such that $\varphi(u) - C|u|^2$ is concave). Moreover, the interpolation $r_t := (1-t)i + tr$ between r and the identity i on Λ is such that $(r_t)_\# \rho$ has a Lebesgue density for all $t \in [0,1]$ (which can e.g. be shown by a modification of the proof of Proposition 9.3.9. in [1]).

Now, assume that f satisfies the McCann condition for geodesic convexity (70), that $\mathcal{F}(\rho), \mathcal{F}(\hat\rho) < \infty$, and that $L_f(\rho) \in W^{1,1}(\Lambda;\mathbb{R})$. Then

$$\int_\Lambda \nabla[L_f(\rho)]\cdot(r-i)\,\mathrm{d}u \leq -\int_\Lambda L_f(\rho)\,\mathrm{tr}\,\tilde\nabla(r-i)\,\mathrm{d}u = \lim_{t\searrow 0}\frac{\mathcal{F}((r_t)_\#\rho) - \mathcal{F}(\rho)}{t} < \infty,$$

where $\tilde\nabla r$ denotes the approximate derivative (see Definition 5.5.1 in [1]) and i is the identity on Λ. This result can be obtained from a modification of the proofs of Lemma 10.4.4 and Lemma 10.4.5 in [1].

For a λ-convex functional $\mathcal{G}$, the Fréchet-subdifferential $\partial\mathcal{G}(\rho)$ at $\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ with $\int_\Lambda \rho\,\mathrm{d}u > 0$ consists of all vectors $\zeta \in \mathcal{L}^2_\rho(\Lambda;\mathbb{R}^d) := \{\zeta\colon \Lambda \to \mathbb{R}^d : \|\zeta\|_\rho < \infty\}$ such that for all $\hat\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ with $\int_\Lambda \rho\,\mathrm{d}u = \int_\Lambda \hat\rho\,\mathrm{d}u$

$$\mathcal{G}(\hat\rho) - \mathcal{G}(\rho) \geq \int_\Lambda \zeta\cdot(r-i)\rho\,\mathrm{d}u + \frac{\lambda}{2}W_2^2(\rho,\hat\rho), \tag{73}$$

where r is the optimal transport map from ρ to $\hat\rho$ (see Equation (10.1.7) in [1]).

Lemma 4.5 (Slope and subdifferential, cf. Theorem 10.4.6 in [1])**.** *Assume that f satisfies the McCann condition for geodesic convexity* (70)*. For $\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ with $\int_\Lambda \rho\,\mathrm{d}u > 0$ and $\mathcal{F}(\rho) < \infty$ the following statements are equivalent.*

1. *The Fréchet-subdifferential* (73) *is non-empty, $\partial\mathcal{F}^V(\rho) \neq \emptyset$.*
2. *The metric derivative at ρ is finite,*

$$|\partial\mathcal{F}^V|(\rho) := \limsup_{W_2(\rho,\hat\rho)\to 0} \frac{(\mathcal{F}^V(\rho) - \mathcal{F}^V(\hat\rho))^+}{W_2(\rho,\hat\rho)} < \infty.$$

3. *$L_f(\rho) \in W^{1,1}_{\mathrm{loc}}(\Lambda;\mathbb{R})$ with $\nabla[L_f(\rho)] + \rho\nabla V = \rho w$ for some $w \in \mathcal{L}^2_{\nabla,\rho}(\Lambda;\mathbb{R}^d)$.*

If either of the above holds we have $w \in \partial\mathcal{F}(\rho)$ *and* $\|w\|_\rho = |\partial\mathcal{F}|(\rho)$. *Moreover, if the additional assumption* (71) *holds, then the above conditions are also equivalent to*

4. $\phi(\rho) \in W^{1,1}_{\mathrm{loc}}(\Lambda;\mathbb{R})$ *with* $\nabla[\phi(\rho)] + \chi(\rho)\nabla V = \chi(\rho)w$ *for some* $w \in \mathcal{L}^2_{\nabla,\chi(\rho)}(\Lambda;\mathbb{R}^d)$.

Proof. The equivalence between 1 and 2 holds since (by Lemma 10.1.5 in [1]) the metric slope for (regular and thus in particular) λ-convex functionals is given by

$$|\partial\mathcal{F}|(\rho) = \min\{\|\zeta\|_\rho : \zeta \in \partial\mathcal{F}(\rho)\}. \tag{74}$$

We next show that 2 implies 3. The result follows from a standard calculation, cf. e.g. the proof of Lemma 3.5 in [26]. Consider a smooth function $\xi \in C^\infty_c(\Lambda;\mathbb{R})$. We define the flow associated to $\nabla\xi$ as the unique solution $X(t,u)$ to $\dot X(t,u) = \nabla\xi(X(t,u)), \quad X(0,u) = u$ for $u \in \Lambda$ and $t \in (0,1)$. For $\rho^\xi_t := X(t,\cdot)_\#\rho$ we have (cf. (3.32) in [26])

$$W_2^2(\rho,\rho^\xi_t) \le t\int_0^t \|\nabla\xi\|^2_{\rho^\xi_s}\,\mathrm{d}s = t^2(\|\nabla\xi\|^2_\rho + o(1)). \tag{75}$$

Similar to (3.35) and (3.36) in [26] one finds

$$\lim_{t\to0}\frac{\mathcal{F}(\rho^\xi_t)-\mathcal{F}(\rho)}{t} = \int_\Lambda \nabla[L_f(\rho)]\cdot\nabla\xi\,\mathrm{d}u \quad \text{and} \quad \lim_{t\to0}\frac{\mathcal{V}(\rho^\xi_t)-\mathcal{V}(\rho)}{t} = \int_\Lambda \rho\nabla V\cdot\nabla\xi\,\mathrm{d}u. \tag{76}$$

Using (75) and $\mathcal{F}^V = \mathcal{F}+\mathcal{V}$ we obtain (cf. (3.33) in [26])

$$|\partial\mathcal{F}^V|(\rho) \ge \frac{1}{\|\nabla\xi\|_\rho}\lim_{t\to0}\frac{\mathcal{F}^V(\rho^\xi_t)-\mathcal{F}^V(\rho)}{t} = \frac{1}{\|\nabla\xi\|_\rho}\int_\Lambda(\nabla[L_f(\rho)]+\rho\nabla V)\cdot\nabla\xi\,\mathrm{d}u.$$

Similar to the discussion at the beginning of Subsection 4.1, $|\partial\mathcal{F}^V|(\rho) < \infty$ implies that the linear operator $v \mapsto \int_\Lambda(\nabla[L_f(\rho)]+\rho\nabla V)\cdot v\,\mathrm{d}u$ from $\mathcal{L}^2_{\nabla,\rho}(\Lambda;\mathbb{R}^d)$ to $\mathbb{R}$ is bounded, such that Riesz' representation theorem implies the existence of $w \in \mathcal{L}^2_{\nabla,\rho}(\Lambda;\mathbb{R}^d)$ for which $\nabla[L_f(\rho)]+\rho\nabla V = \rho w$, such that $L_f(\rho) \in W^{1,1}_{\mathrm{loc}}(\Lambda;\mathbb{R})$. In particular $|\partial\mathcal{F}^V|(\rho) \ge \|w\|_\rho$.

For the implication 3 to 2 consider any $\hat\rho \in \mathcal{L}^1(\Lambda;[0,\infty))$ with $\int_\Lambda \rho\,\mathrm{d}u = \int_\Lambda\hat\rho\,\mathrm{d}u$ and $\mathcal{F}(\hat\rho) < \infty$. Then

$$\mathcal{F}(\hat\rho)-\mathcal{F}(\rho) \ge \lim_{t\to0}\frac{\mathcal{F}((r_t)_\#\rho)-\mathcal{F}(\rho)}{t} \ge \int_\Lambda \nabla[L_F(\rho)]\cdot(r-i)\,\mathrm{d}u,$$

where the fist inequality follows from the monotonicity of the difference quotient (see Equation (10.4.24) in [1]). The λ-convexity of $\mathcal{V}$ yields (cf. (69))

$$\mathcal{V}(\hat\rho)-\mathcal{V}(\rho) \ge \lim_{t\to0}\frac{\mathcal{V}((r_t)_\#\rho)-\mathcal{V}(\rho)}{t} + \frac{\lambda}{2}W_2^2(\rho,\hat\rho) = \int_\Lambda \rho\nabla V\cdot(r-i)\,\mathrm{d}u + \frac{\lambda}{2}W_2^2(\rho,\hat\rho).$$

This implies that $w = (\nabla[L_F(\rho)]/\rho + \nabla V) \in \partial\mathcal{F}^V(\rho)$ and thus $|\partial\mathcal{F}^V|(\rho) \le \|w\|_\rho < \infty$ by eqn. (74).

The equivalence between 3 and 4 can be seen as follows: Recall that $C_*L'_f(a) \le \phi'(a) \le C_{\mathrm{Lip}}L'_f(a)$ and also $C_*L_f(a) \le \phi(a) \le C_{\mathrm{Lip}}L_f(a)$. With the same argument as in the proof of Proposition 4.4 we obtain that the chain rule holds as in (66), i.e. $L'_f(\rho)\nabla\rho = \nabla[L_f(\rho)]$ and $\phi'(\rho)\nabla\rho = \nabla[\phi(\rho)]$, such that $C_*\|\nabla[L_f(\rho)]\| \le \|\nabla[\phi(\rho)]\| \le C_{\mathrm{Lip}}\|\nabla[L_f(\rho)]\|$. This proves that $\phi(\rho) \in W^{1,1}(\Lambda;\mathbb{R})$ if and only if $L_f(\rho) \in W^{1,1}(\Lambda;\mathbb{R})$. Moreover $w = \nabla[L_f(\rho)]/\rho = \nabla[\phi(\rho)]/\chi(\rho)$. ∎

Finally, we can outline a proof for Theorem 3.2, which follows ideas from [1, 26]. Since we work on the torus $\Lambda = \mathbb{T}^d$ (rather than $\mathbb{R}^d$), we sketch the argument.

Sketch of the proof of Theorem 3.2. Since $\mathbb{A}$ is finite and the assumptions of Section 2.4.2 are valid

Propositions 4.1 and 4.3 and 4.4 hold. Moreover, since f satisfies the McCann condition (70) and also the assumption (71) on χ' holds we can apply Lemma 4.5. Combining all these results we have that the map $t \mapsto |\rho_t'||\partial \mathcal{F}^V|(\rho_t)$ is in $\mathcal{L}^1_{\mathrm{loc}}([0,T];\mathbb{R})$. This then implies that $t \mapsto \mathcal{F}^V(\rho_t)$ is locally absolutely continuous (see e.g. Lemma 3.4 in [26]), with a.e. derivative

$$\frac{d}{dt}\mathcal{F}^V(\rho_t) = -\langle v_t, w_t\rangle_{\chi(\rho_t)} = -\langle \dot\rho_t, \Delta(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\rangle_{-1,\chi(\rho_t)},$$

which implies the chain rule (39). □

5 Proofs and Supplementary Content

For nearest neighbour transitions, the following proposition yields a special representation for symmetric summands.

Proposition 5.1. *Let $A_{\eta,\eta'}$ be a symmetric function (such that $A_{\eta,\eta'} = A_{\eta',\eta}$) with $A_{\eta,\eta} = 0$ and $A_{\eta,\eta^{i,j}} = 0$ whenever $|i-j| \neq 1$. If either $\sum_{\eta,\eta'\in\Omega_L} |A_{\eta,\eta'}| < \infty$ or $A_{\eta,\eta'} \geq 0$ for all $\eta,\eta' \in \Omega_L$, then*

$$\sum_{\eta,\eta'\in\Omega_L} A_{\eta,\eta'} = 2\sum_{i\in\mathbb{T}^d_L}\sum_{k=1}^{d}\sum_{\eta\in\Omega_L} A_{\eta,\eta^{i,i+e_k}}\mathbf{1}_{\{\eta(i)>0\}}. \tag{77}$$

Proof. Note that by definition $\sum_{\eta,\eta'\in\Omega_L} A_{\eta,\eta'} = \sum_{i\in\mathbb{T}^d_L}\sum_{k=1}^{d}\sum_{\eta\in\Omega_L}\big(A_{\eta,\eta^{i,i+e_k}} + A_{\eta,\eta^{i,i-e_k}}\big)\mathbf{1}_{\{\eta(i)>0\}}$. Using symmetry, the second summand is equal to $A_{\eta^{i,i-e_k},\eta}$, such that first replacing the configuration η with $\eta^{i-e_k,i}$ before replacing the index i with $i+e_k$ yields (77). □

Following [22] Chapter 5, we define for $\epsilon > 0$ the approximation of the identity $\iota_\epsilon := (2\epsilon)^{-d}\mathbf{1}_{[-\epsilon,\epsilon)^d}(\cdot)$. Recall that the convolution of a measure $\pi \in \mathcal{M}_+(\Lambda)$ with a function $f \in \mathcal{L}^1(\Lambda;\mathbb{R})$ is defined as $[\pi * f](u) := \int_\Lambda f(u'-u)\,\pi(\mathrm{d}u')$. The convolution of ι_ϵ with the empirical measure (18) is the function

$$[\Theta_L(\eta) * \iota_\epsilon](u) = (2\epsilon L)^{-d}\sum_{i\in\mathbb{T}^d_L}\mathbf{1}_{[\frac{2i-1}{2L},\frac{2i+1}{2L})^d}(u)\sum_{j:|i-j|\leq\lfloor\epsilon L\rfloor}\eta(j), \tag{78}$$

which is piecewise constant on $\{[\frac{2i-1}{2L},\frac{2i+1}{2L})^d\}_{i\in\mathbb{T}^d_L}$. This allows us to represent the averaged particle density as a function of the empirical distribution, i.e.

$$[\Theta_L(\eta) * \iota_\epsilon](i/L) = \Big(\frac{2\lfloor\epsilon L\rfloor+1}{2\epsilon L}\Big)^d \eta^{\lfloor\epsilon L\rfloor}(i).$$

For $\pi(\mathrm{d}u) = \rho(u)\,\mathrm{d}u$ the convolution yields $[\pi*\iota_\epsilon](u) = (2\epsilon)^{-d}\int_{[u-\epsilon,u+\epsilon)^d}\rho(u')\,\mathrm{d}u'$. Since $\lim_{\epsilon\to0}[\pi * \iota_\epsilon](u) = \rho(u)$ for almost all $u \in \Lambda$, we define $[\pi*\iota_0](u) := \rho(u)$.

5.1 Proofs of the Statements in Section 3.1

Proof of Proposition 3.1. Recall that $(\mu^L_t)_{t\in[0,T]}$ is finitely supported in the sense that the set $\mathcal{N}_0 := \{\eta \in \Omega_L | \mu^L_t(\eta) > 0 \text{ for some } t \in [0,T]\}$ is finite. Since r^L_t consists of nearest neighbour transitions, also the set $\mathcal{N}_1 := \{(\eta,\eta') \in \Omega_L\times\Omega_L | \mu^L_t(\eta)(r^L_t)_{\eta,\eta'} > 0 \text{ or } \mu^L_t(\eta')(r^L_t)_{\eta',\eta} > 0 \text{ for some } t\in[0,T]\}$ is

finite. Thus the left hand side of (36) is equal to

$$\sum_{\eta\in\mathcal{N}_0}\left[\mu^L_{t_2}(\eta)\log\left(\frac{\mu^L_{t_2}(\eta)}{\nu_\alpha(\eta)}\right)-\mu^L_{t_1}(\eta)\log\left(\frac{\mu^L_{t_1}(\eta)}{\nu_\alpha(\eta)}\right)\right]+\sum_{\eta\in\mathcal{N}_0}\sum_{i\in\mathbb{T}^d_L}\left(\mu^L_{t_2}(\eta)\eta(i)\tilde{V}_{t_2}(\tfrac{i}{L})-\mu^L_{t_1}(\eta)\eta(i)\tilde{V}_{t_1}(\tfrac{i}{L})\right)$$
$$+\log\left(\sum_{\eta\in\Omega_L}\nu_\alpha(\eta)\mathrm{e}^{-\sum_{i\in\mathbb{T}^d_L}\tilde{V}_{t_2}(i/L)\eta(i)}\right)-\log\left(\sum_{\eta\in\Omega_L}\nu_\alpha(\eta)\mathrm{e}^{-\sum_{i\in\mathbb{T}^d_L}\tilde{V}_{t_1}(i/L)\eta(i)}\right).$$

Similar to Theorem 9.2 of Appendix 1 in [22], one then shows using (3) that the latter is equal to

$$\sum_{\eta\in\mathcal{N}_0}\int_{t_1}^{t_2}\frac{\mathrm{d}}{\mathrm{d}t}\left[\mu^L_t(\eta)\log\left(\frac{\mu^L_t(\eta)}{\nu_\alpha(\eta)}\right)\right]\mathrm{d}t+\sum_{\eta\in\mathcal{N}_0}\sum_{i\in\mathbb{T}^d_L}\int_{t_1}^{t_2}\frac{\mathrm{d}}{\mathrm{d}t}\left[\mu^L_t(\eta)\eta(i)\tilde{V}_t(\tfrac{i}{L})\right]\mathrm{d}t$$
$$-\int_{t_1}^{t_2}\sum_{\eta\in\Omega_L}\nu_\alpha^{\tilde{V}_t}(\eta)\sum_{i\in\mathbb{T}^d_L}\eta(i)\,\partial_t\tilde{V}_t(\tfrac{i}{L})\,\mathrm{d}t.$$

A straightforward calculation (using $\partial_t\mu^L_t(\eta)=-\operatorname{div}\jmath^L_t(\eta)$, the fact that the transition rates r^L_t are bounded, and the fact that μ^L_t is supported on a finite number of configurations) allows to show that

$$\mathcal{F}^{\tilde{V}_{t_2}}_{L,\alpha}(\mu^L_{t_2})-\mathcal{F}^{\tilde{V}_{t_1}}_{L,\alpha}(\mu^L_{t_1})=-\sum_{\eta\in\mathcal{N}_0}\int_{t_1}^{t_2}\operatorname{div}\jmath^L_t(\eta)\left(\log\left(\frac{\mu^L_t(\eta)}{\nu_\alpha(\eta)}\right)+1\right)\mathrm{d}t$$
$$-\sum_{\eta\in\mathcal{N}_0}\int_{t_1}^{t_2}\operatorname{div}\jmath^L_t(\eta)\sum_{i\in\mathbb{T}^d_L}\eta(i)\,\tilde{V}_t(\tfrac{i}{L})\,\mathrm{d}t+\int_{t_1}^{t_2}\sum_{\eta\in\Omega_L}\left(\mu^L_t(\eta)-\nu_\alpha^{\tilde{V}_t}(\eta)\right)\sum_{i\in\mathbb{T}^d_L}\eta(i)\,\partial_t\tilde{V}_t(\tfrac{i}{L})\,\mathrm{d}t. \quad (79)$$

Using once more the boundedness of the nearest neighbour transition rates and that μ_0 is supported on finitely many configurations, we can show, employing the bound $\log(\mu^L_t(\eta)/\nu_\alpha(\eta))\le|\log(\nu_\alpha(\eta))|$, that

$$\int_0^T\sum_{\eta,\eta'\in\Omega_L}\left|(\jmath^L_t)_{\eta,\eta'}\log\left(\frac{\mu^L_t(\eta)}{\nu_\alpha(\eta)}\right)\right|\mathrm{d}t\le\int_0^T\sum_{(\eta,\eta')\in\mathcal{N}_1}\left(\mu^L_t(\eta)(r^L_t)_{\eta,\eta'}+\mu^L_t(\eta')(r^L_t)_{\eta',\eta}\right)|\log(\nu_\alpha(\eta))|\mathrm{d}t<\infty.$$

The latter allows us to combine the first two summands on the right hand side of (79), which are equal to $-\sum_{\eta\in\Omega_L}\operatorname{div}\jmath^L_t(\eta)\log(\mu^L_t(\eta)/\nu_\alpha^{\tilde{V}_t}(\eta))=-\langle\jmath^L_t,F^{\tilde{V}_t}(\mu^L_t)\rangle_L$, where the last identity follows by a summation by parts (cf. Equation (15) in [21]). This finishes the proof. □

The proof of Theorem 3.3 relies on an auxiliary statement of independent interest, which we prove first. The result gives sufficient conditions for local equilibration.

Lemma 5.2. *Consider $(P_L)_{L\in\mathbb{N}}$ from Section 2.2.1 with associated density $(\mu^L_t)_{t\in[0,T]}$. Assume there exists $\tilde{V}\in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ such that the inequalities*

$$\limsup_{L\to\infty}\frac{1}{L^d}\int_0^T\mathcal{F}^{\tilde{V}_t}_{L,\alpha}(\mu^L_t)\,\mathrm{d}t<\infty \quad (80)$$

and

$$\limsup_{L\to\infty}\frac{1}{L^d}\int_0^T\Psi^\star_L\big(\mu^L_t,F^{\tilde{V}_t}(\mu^L_t)\big)\,\mathrm{d}t<\infty \quad (81)$$

are satisfied. Then $(\mu^L_{[0,T]})_{L\in\mathbb{N}}$ (where again $\mu^L_{[0,T]}:=\frac{1}{T}\int_0^T\mu^L_t\,\mathrm{d}t$) is in the class considered by the replacement lemma (33). In particular (31) and (32) are satisfied for $(\mu^L_{[0,T]})_{L\in\mathbb{N}}$. Moreover, these assumptions are independent of the choices of $\tilde{V}$ and α: We can replace $\tilde{V}$ with $\tilde{V}+\tilde{H}$ for some $\tilde{H}\in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ and also replace α with $\alpha'\in(0,N_{\max})$ in (80) arbitrary. Then (80) and (81)

are satisfied for $\tilde{V}$ and α if and only if they are satisfied for $\tilde{V}+\tilde{H}$ and α'.

Proof. The bound (80) for $\tilde{V}+\tilde{H}$ and α' follows similar to Remark 1.2 on page 70 of [22]. For (81) note that the basic estimate $\cosh(x+y) \le \cosh(x)\mathrm{e}^{|y|}$ combined with (34) yields

$$\frac{1}{L^d}\int_0^T \Psi_L^\star\big(\mu_t^L, F_\alpha^{\tilde{V}_t+\tilde{H}_t}(\mu_t^L)\big)\,\mathrm{d}t \le \frac{C_{\tilde{H}}}{L^d}\int_0^T \Psi_L^\star\big(\mu_t^L, F^{\tilde{V}_t}(\mu_t^L)\big)\,\mathrm{d}t + 2(C_{\tilde{H}}-1)TC_{\hat{\chi}} \tag{82}$$

for some $C_{\tilde{H}} > 0$ that only depends on H. We thus can restrict to the special case $\tilde{V}_t = 0$. The two bounds needed for the replacement lemma (33) then follow from convexity, i.e. $\mathcal{F}^0_{L,\alpha}(\mu^L_{[0,T]}) \le \frac{1}{T}\int_0^T \mathcal{F}^0_{L,\alpha}(\mu_t^L)\,\mathrm{d}t$ and $\Psi_L^\star\big(\mu^L_{[0,T]}, F^0(\mu^L_{[0,T]})\big) \le \frac{1}{T}\int_0^T \Psi_L^\star\big(\mu_t^L, F^0(\mu_t^L)\big)\,\mathrm{d}t$ (cf. the discussion in Chapter 5.3 near equation (3.1) on page 81 in [22]). □

With this result at hand, we can turn to the proof of Theorem 3.3.

Proof of Theorem 3.3. Since the relative entropy is non-negative, we obtain with a modification of (37) to the time interval $[t,T]$ (for each $t \in [0,T]$) that

$$\begin{aligned}\mathcal{F}^{\tilde{V}_t}_{L,\alpha}(\mu_t^L) &\le \mathcal{F}^{\tilde{V}_T}_{L,\alpha}(\mu_T^L) + \int_t^T \Psi_L(\mu_s^L, \jmath_s^L)\,\mathrm{d}s + \int_t^T \Psi_L^\star\big(\mu_s^L, F_\alpha^{\tilde{V}_s}(\mu_s^L)\big)\,\mathrm{d}s \\ &\qquad - \int_t^T \sum_{i\in\mathbb{T}_L^d} \big(\hat{\rho}_i(\mu_s^L) - \bar{\rho}_{\alpha,\tilde{V}_s}(i)\big)\,\partial_s \tilde{V}_s(\tfrac{i}{L})\,\mathrm{d}s \\ &\le \mathbb{A}_L^{\tilde{V}}\big(Q_L\big) + \mathcal{F}^{\tilde{V}_0}_{L,\alpha}(\mu_0^L) + C_{\tilde{V}}\bigg(TL^dC_{\mathrm{tot}} + \int_0^T \sum_{i\in\mathbb{T}_L^d} \bar{\rho}_{\alpha,\tilde{V}_t}(i/L)\,\mathrm{d}t\bigg),\end{aligned} \tag{83}$$

where $C_{\tilde{V}}$ is a constant that only depends on $\tilde{V}$. Thus

$$\begin{aligned}\limsup_{L\to\infty}\frac{1}{L^d}\int_0^T \mathcal{F}^{\tilde{V}_t}_{L,\alpha}(\mu_t^L)\,\mathrm{d}t &\le \limsup_{L\to\infty}\frac{T}{L^d}\mathbb{A}_L^{\tilde{V}}\big(Q_L\big) + \limsup_{L\to\infty}\frac{T}{L^d}\mathcal{F}^{\tilde{V}_0}_{L,\alpha}(\mu_0^L) \\ &\qquad + T^2C_{\tilde{V}}C_{\mathrm{tot}} + TC_{\tilde{V}}\int_0^T\int_\Lambda \bar{\rho}_{\alpha,\tilde{V}_t}(u)\,\mathrm{d}u\,\mathrm{d}t < \infty.\end{aligned} \tag{84}$$

The second inequality follows from a similar estimate to (83): Consider the second inequality in (83) for $t=0$ and drop the term $\mathcal{F}^V_{L,\alpha}(\mu_T^L) + \int_0^T \Psi_L(\mu_t^L, \jmath_t^L)\,\mathrm{d}t \ge 0$. Then

$$\int_0^T \Psi_L^\star\big(\mu_t^L, F^{\tilde{V}_t}(\mu_t^L)\big)\,\mathrm{d}t \le \mathbb{A}_L^{\tilde{V}}\big(Q_L\big) + \mathcal{F}^{\tilde{V}_0}_{L,\alpha}(\mu_0^L) + 2C_{\tilde{V}}\bigg(TL^dC_{\mathrm{tot}} + \int_0^T \sum_{i\in\mathbb{T}_L^d} \bar{\rho}_{\alpha,\tilde{V}_t}(i/L)\,\mathrm{d}t\bigg)$$

and we can conclude as in (84). We then apply Lemma 5.2 to obtain that the equations (31) and (32) are satisfied for $(\mu^L_{[0,T]})_{L\in\mathbb{N}}$. The independence of V, $\tilde{V}$ and α follows from the considerations in Lemma 5.2. □

5.2 Proofs of Liminf Inequalities

This section is devoted to the proof of the liminf inequalities in the proof of Theorem 3.4. Many of the ideas of the following proofs are borrowed from the entropy method developed in [20]. We here follow the presentation of this method in Chapter 5 of the book by Kipnis and Landim [22]. The results we want to prove are of the form $\liminf_{L\to\infty} B_L \ge B_*$. The general strategy involves replacing B_L by

some (possibly ϵ dependent) C_L^ϵ and to show that

$$\liminf_{\epsilon\to 0}\liminf_{L\to\infty} C_L^\epsilon \geq B_* \qquad \text{and} \qquad \limsup_{\epsilon\to 0}\limsup_{L\to\infty} |B_L - C_L^\epsilon| = 0.$$

5.2.1 Bounds for Ψ_L and $\Psi_L^\star$

In order to achieve the projection to the physical domain anticipated in Section 2.1 we consider functions which are linear in η. For this we fix a function $G \in C^1(\Lambda;\mathbb{R})$ and define $\tilde{G}_L\colon \Omega_L \to \mathbb{R}$ by $\tilde{G}_L(\eta) := L^d\langle \Theta_L(\eta), G\rangle = \sum_{i\in\mathbb{T}_L^d} G(i/L)\eta(i)$, for which the discrete derivative satisfies the identity $\nabla^{\eta,\eta^{i,i+e_k}}\tilde{G}_L = \nabla^{i,i+e_k}G(\cdot/L)$. Note that this last identity allows us to reduce the dependence on the configuration space to a dependence on the physical domain. Choosing the 'force' $F = \nabla\tilde{G}_L$, we obtain with Proposition 5.1 (since all summands are non-negative) that

$$\Psi_L^\star(\mu, \nabla\tilde{G}_L) = 2\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{a}_{i,i+e_k}(\mu)L^2\Big[\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G(\cdot/L)\big) - 1\Big] \tag{85}$$

and similar, for the current $j^G_{\eta,\eta'} = a_{\eta,\eta'}(\mu)\sinh\big(\frac{1}{2}\nabla^{\eta,\eta'}\tilde{G}_L\big)$ associated to the above force (cf. [21])

$$\begin{aligned}\Psi_L(\mu, j^G) = 2\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{a}_{i,i+e_k}(\mu)L^2\Big[&\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G(\cdot/L)\big)\tfrac{1}{2}\nabla^{i,i+e_k}G(\cdot/L)\\ &- \Big(\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G(\cdot/L)\big) - 1\Big)\Big].\end{aligned} \tag{86}$$

We next derive upper bounds for (85) and (86) and a lower bound for $\Psi_L^\star(\mu, F^V(\mu))$.

Proposition 5.3 (Upper bounds for Ψ_L and $\Psi_L^\star$)**.** *Let μ be a measure on Ω_L. Further let $f_{\eta,\eta'} := \nabla^{\eta,\eta'}\tilde{G}_L$ for some $G\colon \Lambda\to\mathbb{R}$ and $j^G_{\eta,\eta'} := a_{\eta,\eta'}(\mu)\sinh\big(\frac{1}{2}\nabla^{\eta,\eta'}\tilde{G}_L\big)$. Then*

$$\Psi_L^\star(\mu, \nabla\tilde{G}_L) \leq \Psi_L(\mu, j^G) \leq \frac{1}{2}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}^0_{i,i+e_k}(\mu)\big[2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G(\cdot/L)\big)\big]^2. \tag{87}$$

Proof. The proof follows from the basic inequalities $\cosh(x)-1 \leq x\sinh(x)-(\cosh(x)-1) \leq \frac{1}{2}\sinh(x)^2$ applied to (85) and (86), together with the inequality $\hat{a}_{i,i+e_k}(\mu) \leq 2\hat{\chi}^0_{i,i+e_k}(\mu)$ stated below (17). □

Proposition 5.4 (Lower bound for $\Psi_L^\star$)**.** *Let μ be a measure on Ω_L, $\alpha \in (0, N_{\max})$ and $V \in C^2(\Lambda;\mathbb{R})$. Then, for any $G\colon\Lambda\to\mathbb{R}$ we have the following lower bound on $\Psi_L^\star\big(\mu, F^V(\mu)\big)$ uniform in α*

$$\Psi_L^\star\big(\mu, F^V(\mu)\big) \geq \sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d\Big[\big(L\hat{j}^V_{i,i+e_k}(\mu)\big)\big(L\nabla^{i,i+e_k}G(\cdot/L)\big) - \frac{1}{2}\hat{\chi}^V_{i,i+e_k}(\mu)\big[L\nabla^{i,i+e_k}G(\cdot/L)\big]^2\Big]. \tag{88}$$

Proof. We use the notation $\rho := \mu/\nu_\alpha^V$ (s.t. ρ is the density of μ with respect to ν_α^V) and $q_{\eta,\eta'} := \nu_\alpha^V(\eta)r^V_{\eta,\eta'}$, such that the relation $q_{\eta,\eta'} = q_{\eta',\eta}$ (detailed balance) holds. Then $F^V_{\eta,\eta'}(\mu) = -\nabla^{\eta,\eta'}\log\rho$ and $a_{\eta,\eta'}(\mu) = 2\sqrt{\rho(\eta)q_{\eta,\eta'}\rho(\eta')q_{\eta',\eta}}$. Further, $a_{\eta,\eta'}(\mu)[\cosh(\frac{1}{2}F^V_{\eta,\eta'}(\mu)) - 1] = \sqrt{q_{\eta,\eta'}q_{\eta',\eta}}(\sqrt{\rho(\eta)} - \sqrt{\rho(\eta')})^2$. Using the representation in Proposition 5.1 and $q_{\eta,\eta'} = \sqrt{q_{\eta,\eta'}q_{\eta',\eta}} = q_{\eta',\eta}$, we obtain

$$\Psi_L^\star\big(\mu, F^V(\mu)\big) = \sum_{\eta\in\Omega_L}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d 2q_{\eta,\eta^{i,i+e_k}}\Big(\sqrt{\rho(\eta)} - \sqrt{\rho(\eta^{i,i+e_k})}\Big)^2.$$

Define $H_{\eta,\eta'} = \frac{1}{4}\big(\sqrt{\rho(\eta)}+\sqrt{\rho(\eta')}\big)\nabla^{\eta,\eta'}\tilde{G}_L$. Using $\nabla^{\eta,\eta^{i,i+e_k}}\tilde{G}_L = \nabla^{i,i+e_k}G(\cdot/L)$ one easily establishes

$$\begin{aligned}2\Big(\sqrt{\rho(\eta)}-\sqrt{\rho(\eta^{i,i+e_k})}\Big)^2 &\geq 4\Big(\sqrt{\rho(\eta)}-\sqrt{\rho(\eta^{i,i+e_k})}\Big)H_{\eta,\eta^{i,i+e_k}} - 2H^2_{\eta,\eta^{i,i+e_k}}\\ &= \big(\rho(\eta)-\rho(\eta^{i,i+e_k})\big)\nabla^{i,i+e_k}G(\cdot/L) - \frac{1}{8}\big(\sqrt{\rho(\eta)}+\sqrt{\rho(\eta^{i,i+e_k})}\big)^2(\nabla^{i,i+e_k}G(\cdot/L))^2.\end{aligned}$$

Using $q_{\eta,\eta'} = q_{\eta',\eta}$, the inequality $\frac{1}{2}(x+y)^2 \leq x^2+y^2$, and $\mu(\eta)r^V_{\eta,\eta'} = \rho(\eta)q_{\eta,\eta'}$ thus allows to bound $2q_{\eta,\eta^{i,i+e_k}}(\sqrt{\rho(\eta)}-\sqrt{\rho(\eta^{i,i+e_k})})^2$ from below by

$$\begin{aligned}&\big(\mu(\eta)r^V_{\eta,\eta^{i,i+e_k}} - \mu(\eta^{i,i+e_k})r^V_{\eta^{i,i+e_k},\eta}\big)\nabla^{i,i+e_k}G(\cdot/L)\\ &\qquad - \frac{1}{4}\big(\mu(\eta)r^V_{\eta,\eta^{i,i+e_k}} + \mu(\eta^{i,i+e_k})r^V_{\eta^{i,i+e_k},\eta}\big)(\nabla^{i,i+e_k}G(\cdot/L))^2.\end{aligned}$$

Note that $\sum_{\eta\in\Omega_L}\mu(\eta)r^V_{\eta,\eta^{i+e_k,i}} = \sum_{\eta\in\Omega_L}\mu(\eta^{i,i+e_k})r^V_{\eta^{i,i+e_k},\eta}$ implies that

$$\begin{aligned}\Psi^\star_L\big(\mu,F^V(\mu)\big) \geq \sum_{i\in\mathbb{T}^d_L}\sum_{k=1}^d\Big[&\Big(\sum_{\eta\in\Omega_L}\mu(\eta)\big(r^V_{\eta,\eta^{i,i+e_k}} - r^V_{\eta,\eta^{i+e_k,i}}\big)\Big)\nabla^{i,i+e_k}G(\cdot/L)\\ &- \frac{1}{4}\Big(\sum_{\eta\in\Omega_L}\mu(\eta)\big(r^V_{\eta,\eta^{i,i+e_k}} + r^V_{\eta,\eta^{i+e_k,i}}\big)\Big)\big(\nabla^{i,i+e_k}G(\cdot/L)\big)^2\Big],\end{aligned}\tag{89}$$

which coincides by (15) and (17) with the right hand side of (88). □

5.2.2 Asymptotic Lower Bound for the Free Energy

Proposition 5.5. *Let the assumptions of Theorem 3.4 hold and let $t \in [0,T]$ be such that the path $(\pi_t)_{t\in[0,T]}$ is continuous at t. Then*

$$\liminf_{L\to\infty}\frac{1}{L^d}\mathcal{F}^V_{L,\alpha}\big(\mu^L_t\big) \geq \mathcal{F}^V_\alpha(\rho_t).\tag{90}$$

Proof. For each $h \in C(\Lambda;\mathbb{R})$ the entropy inequality (a special case of the Fenchel inequality, see Proposition 8.1 and page 340 in Appendix 1 in [22]) implies

$$\begin{aligned}\frac{1}{L^d}\mathcal{F}^V_{L,\alpha}\big(\mu^L_t\big) &\geq \frac{1}{L^d}\Big[\sum_{\eta\in\Omega_L}\mu^L_t(\eta)\sum_{i\in\mathbb{T}^d_L}h(i/L)\eta(i) - \log\Big(\sum_{\eta\in\Omega_L}\nu^V_\alpha(\eta)\mathrm{e}^{\sum_{i\in\mathbb{T}^d_L}h(i/L)\eta(i)}\Big)\Big]\\ &= \sum_{\eta\in\Omega_L}\mu^L_t(\eta)\langle\Theta_L(\eta),h\rangle - \frac{1}{L^d}\sum_{i\in\mathbb{T}^d_L}\log\Big(\frac{E_{\nu_{\alpha,1}}[\mathrm{e}^{(h(i/L)-V(i/L))\eta(0)}]}{E_{\nu_{\alpha,1}}[\mathrm{e}^{-V(i/L)\eta(0)}]}\Big).\end{aligned}$$

By the assumption of finite moments in (3) the dominated convergence theorem yields that $u \mapsto E_{\nu_{\alpha,1}}[\mathrm{e}^{(h(u)-V(u))\eta(0)}]$ is continuous.

By (7), we can restrict to measures with bounded volume, such that a truncation argument, combined with the weak convergence $Q^L \to Q^* = \delta_{(\pi_t)_{t\in[0,T]}}$ and the continuity of the projection/evaluation at time t implies $\sum_{\eta\in\Omega_L}\mu^L_t(\eta)\langle\Theta_L(\eta),h\rangle = \mathbb{E}_{Q_L}[\langle\pi_t,h\rangle] \to \mathbb{E}_{Q_L}[\langle\pi_t,h\rangle] = \langle\pi_t,h\rangle$. Thus

$$\liminf_{L\to\infty}\frac{1}{L^d}\mathcal{F}^V_{L,\alpha}\big(\mu^L_t\big) \geq \langle\pi_t,h\rangle - \int_\Lambda\log\Big(\frac{E_{\nu_{\alpha,1}}[\mathrm{e}^{(h(u)-V(u))\eta(0)}]}{E_{\nu_{\alpha,1}}[\mathrm{e}^{-V(u)\eta(0)}]}\Big)\,\mathrm{d}u.\tag{91}$$

Taking the supremum with respect to $h \in C(\Lambda;\mathbb{R})$ combined with (21) then finishes the proof. □

5.2.3 Asymptotic Lower Bound for Ψ

The following proofs will depend on uniform continuity of functions (which follows here from continuity and the compactness of the domain Λ (or $[0,T]\times\Lambda$)).

Lemma 5.6. *Under the assumptions of Theorem 3.4, we have for any $G\in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$*

$$\limsup_{\epsilon\to 0}\limsup_{L\to\infty}\Bigg|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}_{i,i+e_k}(\mu_t^L)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2\,\mathrm{d}t \\ -\int_0^T\int_\Lambda \sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](u)\big)|\nabla G_t(u)|^2\,\mathrm{d}u\,\mathrm{d}t\Bigg| = 0. \tag{92}$$

Proof. We first show that without loss of generality we can set $V=0$ for the rates (1). We denote with $\hat{\chi}^V$ the mobility for a smooth potential V and with $\hat{\chi}^0$ the mobility for $V=0$. Note that

$$\Bigg|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d\Big(\hat{\chi}^V_{i,i+e_k}(\mu_t^L)-\hat{\chi}^0_{i,i+e_k}(\mu_t^L)\Big)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2\mathrm{d}t\Bigg| \\ \le\int_0^T\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d\hat{\chi}^0_{i,i+e_k}(\mu_t^L)2\big(\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big)-1\big)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2\mathrm{d}t. \tag{93}$$

Taylor's theorem enables us to find for each $t\in[0,T]$ a number $\xi\in(i/L,(i+e_k)/L)$ for which $L\nabla^{i,i+e_k}G_t(\cdot/L)=\partial_k G_t(\xi)$. Defining $C_G:=\sum_{k=1}^d\sup_{t\in[0,T]}\|\partial_k G_t\|_\infty^2<\infty$ allows us to bound the right hand side of (93) from above by

$$\frac{2C_GT}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d\big(\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big)-1\big)\hat{\chi}^0_{i,i+e_k}\left(\frac{1}{T}\int_0^T\mu_t^L\,\mathrm{d}t\right). \tag{94}$$

Using the uniform continuity of V (on the compact set Λ), we obtain for each $\epsilon>0$ that $|\nabla^{i,i+e_k}V(\cdot/L)|<\epsilon$ as $L\to\infty$ independent of i and e_k, such that (94) is (for L large enough) with (34) bounded by $2C_GC_{\hat{\chi}}T(\cosh(\epsilon/2)-1)$. Thus, taking the limit superior $\epsilon\to 0$ after taking $L\to\infty$ in (94) shows that the left hand side of (93) vanishes. This justifies the replacement of V with $V=0$ in the mobility. We thus drop the indices V and 0 and simply write $\hat{\chi}$ for the mobility with $V=0$.

To prove (92) it is sufficient to show that

$$
\begin{aligned}
&\limsup_{\epsilon\to 0}\limsup_{L\to\infty}\\
&\quad\Big|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\frac{\hat\chi_{i,i+e_k}(\mu_t^L)}{(2\lfloor\epsilon L\rfloor+1)^d}\sum_{|m|\le\lfloor\epsilon L\rfloor}\Big(\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2-\big[\partial_k G_t((i+m)/L)\big]^2\Big)\,\mathrm{d}t\Big|\\
&\quad+\frac{C_G T}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\sum_{\eta\in\Omega_L}\Big(\frac{1}{T}\int_0^T\mu_t^L(\eta)\mathrm{d}t\Big)\Big|\hat\chi^{\lfloor\epsilon L\rfloor}_{i,i+e_k}(\delta_\eta)-\hat\chi_{i,i+e_k}\big(\nu_{\eta^{\lfloor\epsilon L\rfloor}(i)}\big)\Big|\\
&\quad+\frac{C_G T}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{\eta\in\Omega_L}\Big(\frac{1}{T}\int_0^T\mu_t^L(\eta)\mathrm{d}t\Big)\Big|\chi\big(\eta^{\lfloor\epsilon L\rfloor}(i)\big)-\chi\Big(\Big(\frac{2\epsilon L}{2\lfloor\epsilon L\rfloor+1}\Big)^d\eta^{\lfloor\epsilon L\rfloor}(i)\Big)\Big|\\
&\quad+\Big|\int_0^T\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](i/L)\big)|\nabla G_t(i/L)|^2\,\mathrm{d}t\\
&\qquad\qquad-\int_0^T\int_\Lambda\sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](u)\big)|\nabla G_t(u)|^2\,\mathrm{d}u\,\mathrm{d}t\Big|=0.
\end{aligned}
\tag{95}
$$

By uniform continuity of $(\partial_k G_t)^2$ for each $\delta>0$ there exists an $\epsilon>0$ such that $|u-u'|<\epsilon$ implies that $|(\partial_k G_t(u))^2-(\partial_k G_t(u'))^2|<\delta$ uniformly in $t\in[0,T]$. Thus, by (34), the first term in (95) is, for ϵ small enough, bounded by

$$
\int_0^T\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\frac{\hat\chi_{i,i+e_k}(\mu_t^L)}{(2\lfloor\epsilon L\rfloor+1)^d}\sum_{|m|\le\lfloor\epsilon L\rfloor}\Big|\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2-\big[\partial_k G_t((i+m)/L)\big]^2\Big|\,\mathrm{d}t\le T\delta C_{\hat\chi}.
$$

Letting $\delta\to 0$ shows that the first term in (95) vanishes.

The second term is controlled by the local equilibrium assumption (31); the third term vanishes using the Lipschitz continuity of χ and the bound on the expected number of particles: The Lipschitz continuity yields that the third summand in (95) is bounded by

$$
C_G C_{\mathrm{Lip}} T\Big|1-\Big(\frac{2\epsilon L}{2\lfloor\epsilon L\rfloor+1}\Big)^d\Big|\sum_{\eta\in\Omega_L}\Big(\frac{1}{T}\int_0^T\mu_t^L(\eta)\,\mathrm{d}t\Big)\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\eta^{\lfloor\epsilon L\rfloor}(i).
$$

By the conservation of particles, the last expression can be bounded by $C_G C_{\mathrm{Lip}} C_{\mathrm{tot}} T\big|1-\big(\frac{2\epsilon L}{2\lfloor\epsilon L\rfloor+1}\big)^d\big|$, which vanishes as $L\to\infty$.

For the last term in (95) recall that $[\Theta_L(\eta)*\iota_\epsilon](u)$ is piecewise constant on $\{[\frac{2i-1}{2L},\frac{2i+1}{2L})^d\}_{i\in\mathbb{T}_L^d}$ (cf. (78)). The proof thus reduces to establishing a bound for

$$
\int_0^T\sum_{i\in\mathbb{T}_L^d}\sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](i/L)\big)\Big|\int_{[\frac{2i-1}{2L},\frac{2i+1}{2L})^d}\big(|\nabla G_t(i/L)|^2-|\nabla G_t(u)|^2\big)\,\mathrm{d}u\Big|\,\mathrm{d}t,
$$

which is easily obtained, as the the last expression is by the Lipschitz continuity, (7), and (78) bounded above by

$$
C_{\mathrm{Lip}}C_{\mathrm{tot}}(2\epsilon)^{-d}\int_0^T\sum_{i\in\mathbb{T}_L^d}\int_{[\frac{2i-1}{2L},\frac{2i+1}{2L})^d}\big|\nabla G_t(i/L)|^2-|\nabla G_t(u)|^2\big|\,\mathrm{d}u\,\mathrm{d}t,
$$

which converges by the uniform continuity of ∇G to zero for $L\to\infty$. □

Note that the above proof does not depend on the fact that we consider the square gradient of a function G. We can replace the square by the product of two different gradients and immediately

obtain the following results.

Lemma 5.7. *Under the assumptions of Theorem 3.4 we have for any $G, H \in C^1([0,T]\times\Lambda;\mathbb{R})$ that*

$$\limsup_{\epsilon\to 0}\limsup_{L\to\infty}\Big|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}_{i,i+e_k}(\mu_t^L)\big[L\nabla^{i,i+e_k}H_t(\cdot/L)\big]\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]\,\mathrm{d}t - \int_0^T\int_\Lambda \sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](u)\big)\nabla H_t(u)\cdot\nabla G_t(u)\,\mathrm{d}u\,\mathrm{d}t\Big| = 0. \quad (96)$$

Corollary 5.8. *Under the assumptions of Theorem 3.4 we have for any $G \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ that*

$$\limsup_{\epsilon\to 0}\limsup_{L\to\infty}\Big|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}_{i,i+e_k}(\mu_t^L)\big[2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G_t(\cdot/L)\big)\big]^2\,\mathrm{d}t - \int_0^T\int_\Lambda \sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](u)\big)|\nabla G_t(u)|^2\,\mathrm{d}u\,\mathrm{d}t\Big| = 0 \quad (97)$$

and for any $G, H \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$

$$\limsup_{\epsilon\to 0}\limsup_{L\to\infty}\Big|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}_{i,i+e_k}(\mu_t^L)\big[2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G_t(\cdot/L)\big)\big]\big[L\nabla^{i,i+e_k}H_t(\cdot/L)\big]\,\mathrm{d}t - \int_0^T\int_\Lambda \sum_{\eta\in\Omega_L}\mu_t^L(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](u)\big)\nabla G_t(u)\cdot\nabla H_t(u)\,\mathrm{d}u\,\mathrm{d}t\Big| = 0. \quad (98)$$

We now turn to the proof of the lower bound in (51).

Proposition 5.9. *Let the assumptions of Theorem 3.4 hold. Then* (51) *is satisfied.*

Proof. For any $G \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ we have

$$\sum_{\eta\in\Omega_L}\tilde{G}_L(T,\eta)\mu_T^L(\eta) - \sum_{\eta\in\Omega_L}\tilde{G}_L(0,\eta)\mu_0^L(\eta) - \int_0^T\sum_{\eta\in\Omega_L}\partial_t\tilde{G}_L(t,\eta)\mu_t^L(\eta)\,\mathrm{d}t = \int_0^T\langle \jmath_t^L, \nabla\tilde{G}_L(t,\cdot)\rangle_L\,\mathrm{d}t \le \int_0^T \Psi_L(\mu_t^L,\jmath_t^L)\,\mathrm{d}t + \int_0^T \Psi_L^\star(\mu_t^L,\nabla\tilde{G}_L(t,\cdot))\,\mathrm{d}t. \quad (99)$$

Combined with Proposition 5.3 we obtain that $\frac{1}{L^d}\int_0^T \Psi_L(\mu_t^L,\jmath_t^L)\,\mathrm{d}t$ is bounded below by

$$\sum_{\eta\in\Omega_L}\mu_T^L(\eta)\langle\Theta_L(\eta),G_T\rangle - \sum_{\eta\in\Omega_L}\mu_0^L(\eta)\langle\Theta_L(\eta),G_0\rangle - \int_0^T\sum_{\eta\in\Omega_L}\mu_t^L(\eta)\langle\Theta_L(\eta),\partial_t G_t\rangle\,\mathrm{d}t - \frac{1}{2L^d}\int_0^T\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}^0_{i,i+e_k}(\mu_t^L)\big[2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G_t(\cdot/L)\big)\big]^2\,\mathrm{d}t. \quad (100)$$

For $\epsilon>0$ and G fixed we define the function $f^{\epsilon,G}\colon \mathcal{D}([0,T];\mathcal{M}_+(\Lambda))\to\mathbb{R}$ which assigns to a path $(\tilde{\pi}_t)_{t\in[0,T]}$ the value

$$f^{\epsilon,G}((\tilde{\pi}_t)_{t\in[0,T]}) := \langle\tilde{\pi}_T,G_T\rangle - \langle\tilde{\pi}_0,G_0\rangle - \int_0^T\langle\tilde{\pi}_t,\partial_t G_t\rangle\,\mathrm{d}t - \frac{1}{2}\int_0^T\int_\Lambda \chi\big([\tilde{\pi}_t*\iota_\epsilon](u)\big)|\nabla G_t(u)|^2\,\mathrm{d}u\,\mathrm{d}t.$$

By (7), we can restrict $f^{\epsilon,G}$ to measures with bounded volume. In this case $f^{\epsilon,G}$ is continuous and bounded, which follows from dominated convergence using the estimate $\chi\big([\pi_t * \iota_\epsilon](u)\big)|\nabla G_t(u)|^2 \leq C_G C_{\mathrm{Lip}} C_{\mathrm{tot}}/(2\epsilon)^d < \infty$. We can rewrite (100) as

$$\mathbb{E}_{Q_L}\big[f^{\epsilon,G}\big] + \frac{1}{2}\int_0^T \int_\Lambda \sum_{\eta\in\Omega_L} \mu_t^L(\eta)\chi\big([\Theta_L(\eta) * \iota_\epsilon](u)\big)|\nabla G_t|^2 \,\mathrm{d}u\,\mathrm{d}t \\ - \frac{1}{2L^d}\int_0^T \sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}_{i,i+e_k}(\mu_t^L)\big[2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G_t(\cdot/L)\big)\big]^2\,\mathrm{d}t$$

and define the remainder

$$R_L^\epsilon := \frac{1}{2}\Big|\int_0^T \frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d \hat{\chi}_{i,i+e_k}(\mu_t)\big[2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}G_t(\cdot/L)\big)\big]^2\,\mathrm{d}t \\ - \int_0^T\int_\Lambda \sum_{\eta\in\Omega_L}\mu_t(\eta)\chi\big([\Theta_L(\eta)*\iota_\epsilon](u)\big)|\nabla G_t|^2\,\mathrm{d}u\,\mathrm{d}t\Big|$$

to obtain $L^{-d}\int_0^T \Psi_L(\mu_t, j_t)\,\mathrm{d}t \geq \mathbb{E}_{Q_L}\big[f^{\epsilon,G}\big] - R_L^\epsilon$.

Since $f^{\epsilon,G}$ is continuous and bounded, the weak convergence $Q_L \to Q^* = \delta_{(\pi_t)_{t\in[0,T]}}$ implies that $\lim_{L\to\infty}\mathbb{E}_{Q_L}\big[f^{\epsilon,G}\big] = \mathbb{E}_{Q^*}\big[f^{\epsilon,G}\big] = f^{\epsilon,G}((\pi_t)_{t\in[0,T]})$. Furthermore $\limsup_{\epsilon\to0}\limsup_{L\to\infty} R_L^\epsilon = 0$ by Corollary 5.8. Thus $\liminf_{L\to\infty} L^{-d}\int_0^T \Psi_L(\mu_t,j_t)\,\mathrm{d}t \geq \liminf_{\epsilon\to0} f^{\epsilon,G}((\pi_t)_{t\in[0,T]})$.

For $\pi_t(\mathrm{d}u) = \rho_t(u)\,\mathrm{d}u$ the distance $|f^{\epsilon,G}((\pi_t)_{t\in[0,T]}) - f^{0,G}((\pi_t)_{t\in[0,T]})|$ is bounded from above by

$$\frac{C_G}{2}\int_0^T\int_\Lambda \Big|\chi\big([\rho_t*\iota_\epsilon](u)\big) - \chi(\rho_t(u))\Big|\mathrm{d}u\,\mathrm{d}t \leq \frac{C_G C_{\mathrm{Lip}}}{2}\int_0^T\int_\Lambda \big|[\rho_t*\iota_\epsilon](u) - \rho_t(u)\big|\mathrm{d}u\,\mathrm{d}t, \tag{101}$$

which is integrable. The dominated convergence theorem then implies that $f^{\epsilon,G}((\pi_t)_{t\in[0,T]}) \to f^{0,G}((\pi_t)_{t\in[0,T]})$, which proves $\liminf_{L\to\infty} L^{-d}\int_0^T \Psi_L(\mu_t,j_t)\,\mathrm{d}t \geq f^{0,G}((\pi_t)_{t\in[0,T]})$. Taking the supremum over all $G \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ finally yields (51). □

5.2.4 Asymptotic Lower Bound for $\Psi^\star$

The proofs in this section are very similar to the proofs in Section 5.2.3. We will therefore be brief.

Lemma 5.10. *Suppose the assumptions of Theorem 3.4 hold. Then*

$$\limsup_{\epsilon\to0}\limsup_{L\to\infty} \Big|\int_0^T\Big(\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^d\Big[\big(L\hat{\jmath}^V_{i,i+e_k}(\mu_t^L)\big)\big(L\nabla^{i,i+e_k}G_t(\cdot/L)\big) - \frac{1}{2}\hat{\chi}^V_{i,i+e_k}(\mu_t^L)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2\Big] \\ -\mathbb{E}_{Q_L}\Big[\int_\Lambda \phi\big([\pi_t*\iota_\epsilon](u)\big)\Delta G_t\,\mathrm{d}u - \int_\Lambda \chi\big([\pi_t*\iota_\epsilon](u)\big)\nabla V\cdot\nabla G_t\,\mathrm{d}u - \frac{1}{2}\int_\Lambda \chi\big([\pi_t*\iota_\epsilon](u)\big)|\nabla G_t|^2\,\mathrm{d}u\Big]\Big)\mathrm{d}t\Big| = 0. \tag{102}$$

Proof. Note that

$$\hat{\jmath}^V_{i,i+e_k}(\mu) = \hat{\jmath}^0_{i,i+e_k}(\mu)\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big) + \hat{\chi}^0_{i,i+e_k}(\mu)2\sinh\big(-\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big). \tag{103}$$

Using (16) and (103), a discrete integration by parts (i.e. a shift of the index) yields

$$\begin{aligned}
&\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\big(L\hat{\jmath}^V_{i,i+e_k}(\mu)\big)\big(L\nabla^{i,i+e_k}G_t(\cdot/L)\big)-\frac{1}{2}\hat{\chi}_{i,i+e_k}(\mu)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2\\
&=\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\hat{\phi}_i(\mu)L^2\Big[\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big)\nabla^{i,i+e_k}G_t(\cdot/L)-\cosh\big(\tfrac{1}{2}\nabla^{i-e_k,i}V(\cdot/L)\big)\nabla^{i-e_k,i}G_t(\cdot/L)\Big]\\
&\qquad+\hat{\chi}^0_{i,i+e_k}(\mu)2L\sinh\big(-\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big)\big(L\nabla^{i,i+e_k}G_t(\cdot/L)\big)-\frac{1}{2}\hat{\chi}_{i,i+e_k}(\mu)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2.
\end{aligned}$$

Combining this with the expression in (102), it is sufficient to show that

$$\begin{aligned}
\limsup_{\epsilon\to0}\limsup_{L\to\infty}\Bigg|\int_0^T\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\hat{\phi}_i(\mu_t^L)L^2\Big[&\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big)\nabla^{i,i+e_k}G_t(\cdot/L)\\
&-\cosh\big(\tfrac{1}{2}\nabla^{i-e_k,i}V(\cdot/L)\big)\nabla^{i-e_k,i}G_t(\cdot/L)\Big]-\mathbb{E}_{Q_L}\bigg[\int_\Lambda\phi\big([\pi_t*\iota_\epsilon](u)\big)\Delta G_t(u)\,\mathrm{d}u\bigg]\,\mathrm{d}t\Bigg|=0, \quad (104)
\end{aligned}$$

as well as

$$\begin{aligned}
\limsup_{\epsilon\to0}\limsup_{L\to\infty}\Bigg|\int_0^T\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\hat{\chi}^0_{i,i+e_k}(\mu_t^L)2L\sinh\big(\tfrac{1}{2}\nabla^{i,i+e_k}V(\cdot/L)\big)\big(L\nabla^{i,i+e_k}G_t(\cdot/L)\big)&\\
-\mathbb{E}_{Q_L}\bigg[\int_\Lambda\chi\big([\pi_t*\iota_\epsilon](u)\big)\nabla V(u)\cdot\nabla G_t(u)\,\mathrm{d}u\bigg]\,\mathrm{d}t\Bigg|& \quad (105)
\end{aligned}$$

$$+\frac{1}{2}\Bigg|\int_0^T\frac{1}{L^d}\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\hat{\chi}_{i,i+e_k}(\mu_t^L)\big[L\nabla^{i,i+e_k}G_t(\cdot/L)\big]^2-\mathbb{E}_{Q_L}\bigg[\int_\Lambda\chi\big([\pi_t*\iota_\epsilon](u)\big)|\nabla G_t(u)|^2\,\mathrm{d}u\bigg]\,\mathrm{d}t\Bigg|=0.$$

Note that (105) follows from the above considerations (Lemma 5.7 and Corollary 5.8), such that we are only left to prove (104), which can be proven with the same calculations as above (with $\hat{\chi}$ replaced by $\hat{\phi}$ combined with (34) and using (32) instead of (31)). □

Proposition 5.11. *Under the assumptions of Theorem 3.4 the inequality* (52) *holds.*

Proof. We only sketch the proof, which is very similar to the one of Proposition 5.9. For

$$\begin{aligned}
f^{\epsilon,G}((\tilde{\pi}_t)_{t\in[0,T]}):=\int_0^T\int_\Lambda\phi\big([\tilde{\pi}_t*\iota_\epsilon](u)\big)\Delta G_t\,\mathrm{d}u\,\mathrm{d}t-\int_0^T\int_\Lambda\chi\big([\tilde{\pi}_t*\iota_\epsilon](u)\big)\nabla V\cdot\nabla G_t\,\mathrm{d}u\,\mathrm{d}t&\\
-\frac{1}{2}\int_0^T\int_\Lambda\chi\big([\tilde{\pi}_t*\iota_\epsilon](u)\big)|\nabla G_t|^2\,\mathrm{d}u\,\mathrm{d}t.&
\end{aligned}$$

Proposition 5.4 implies that

$$\Psi_L^\star\big(\mu,F^V(\mu)\big)\geq\sum_{i\in\mathbb{T}_L^d}\sum_{k=1}^{d}\Big[\big(L\hat{\jmath}^V_{i,i+e_k}(\mu)\big)\big(L\nabla^{i,i+e_k}G(\cdot/L)\big)-\frac{1}{2}\hat{\chi}^V_{i,i+e_k}(\mu)\big[L\nabla^{i,i+e_k}G(\cdot/L)\big]^2\Big].$$

As in the proof of Proposition 5.9, one obtains $\frac{1}{L^d}\int_0^T\Psi_L^\star(\mu_t^L,F^S(\mu_t^L))\,\mathrm{d}t\geq\mathbb{E}_{Q_L}\big[f^{\epsilon,G}\big]-R_L^\epsilon$, where R_L^ϵ coincides with (102) in Lemma 5.10. The latter implies that $\limsup_{\epsilon\to0}\limsup_{L\to\infty}R_L^\epsilon=0$, such

that again by weak convergence with $\epsilon \to 0$

$$\liminf_{L\to\infty} \frac{1}{L^d} \int_0^T \Psi_L^\star(\mu_t^L, F^S(\mu_t^L))\,\mathrm{d}t \geq f^{0,G}((\pi_t)_{t\in[0,T]}).$$

Taking the supremum with respect to $G \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$ yields (52). □

5.3 Proof of Theorem 3.5

Proof of Theorem 3.5. We extend the proof in [3]. We will skip some details, as they are similar to the above calculations. Let $\tilde{H} \in C^{1,2}([0,T]\times\Lambda;\mathbb{R})$. The log density of $P_L^{V+\tilde{H}}$ with respect to P_L^V (where both measures have the same initial condition μ_0^L) has the explicit representation (cf. [3] and the Appendix in [21])

$$\log \frac{dP_L^{V+\tilde{H}}}{dP_L^V}((\eta_t)_{t\in[0,T]}) = \frac{L^d}{2}\Big[\langle \Theta_L(\eta_T), \tilde{H}_T\rangle - \langle \Theta_L(\eta_0), \tilde{H}_0\rangle - \int_0^T \langle \Theta_L(\eta_t), \partial_t \tilde{H}_t\rangle\,\mathrm{d}t\Big]$$
$$- \int_0^T \sum_{i\in\mathbb{T}_L^d} \sum_{i':|i-i'|=1} \hat{r}^V_{\eta_t,\eta_t^{i,i'}} L^2\big(\mathrm{e}^{-\frac{1}{2}(\tilde{H}_t(i'/L)-\tilde{H}_t(i/L))} - 1\big)\mathrm{d}t.$$

Using $2(ac+bd) = (a-b)(c-d)+(a+b)(c+d)$ we can represent the expression in the last line as

$$\int_0^T \sum_{i\in\mathbb{T}_L^d} \sum_{k=1}^d \Big[L\big(\hat{r}^V_{\eta_t,\eta_t^{i,i+e_k}} - \hat{r}^V_{\eta_t,\eta_t^{i+e_k,i}}\big)\big(L\sinh\big(-\tfrac{1}{2}\nabla^{i,i+e_k}\tilde{H}_t(\tfrac{\cdot}{L})\big)\big)$$
$$+ \big(\hat{r}^V_{\eta_t,\eta_t^{i,i+e_k}} + \hat{r}^V_{\eta_t,\eta_t^{i+e_k,i}}\big)L^2\big(\cosh\big(-\tfrac{1}{2}\nabla^{i,i+e_k}\tilde{H}_t(\tfrac{\cdot}{L})\big) - 1\big)\Big]\mathrm{d}t.$$

Taking the expected value of this expression with respect to P_L^V, in combined with (15) and (17), yields

$$\int_0^T \sum_{i\in\mathbb{T}_L^d} \sum_{k=1}^d \Big[\big(L\hat{\jmath}^V_{i,i+e_k}(\mu_t^L)\big)\big(L\sinh\big(-\tfrac{1}{2}\nabla^{i,i+e_k}\tilde{H}_t(\tfrac{\cdot}{L})\big)\big)$$
$$+ 2\hat{\chi}^V_{i,i+e_k}(\mu_t^L)L^2\big(\cosh\big(\tfrac{1}{2}\nabla^{i,i+e_k}\tilde{H}_t(\tfrac{\cdot}{L})\big) - 1\big)\Big]\mathrm{d}t, \tag{106}$$

which is asymptotically equivalent to

$$\int_0^T \frac{1}{2}\sum_{i\in\mathbb{T}_L^d} \sum_{k=1}^d \Big[-\big(L\hat{\jmath}^V_{i,i+e_k}(\mu_t^L)\big)\big(L\nabla^{i,i+e_k}\tilde{H}_t(\tfrac{\cdot}{L})\big) + \frac{1}{2}\hat{\chi}^V_{i,i+e_k}(\mu_t^L)L^2\big|\nabla^{i,i+e_k}\tilde{H}_t(\tfrac{\cdot}{L})\big|^2\Big]\mathrm{d}t.$$

A result similar to Lemma 5.10 yields

$$\lim_{L\to\infty}\frac{1}{L^d}\mathbb{A}_L^V\big(Q_L^{V+\tilde{H}}\big) = \lim_{\epsilon\to 0}\lim_{L\to\infty}\frac{1}{2}\mathbb{E}_{Q_L}\big[f^{\epsilon,\tilde{H}}\big] = \frac{1}{2}f^{0,\tilde{H}}((\pi_t)_{t\in[0,T]}),$$

where the functional $f^{\epsilon,\tilde{H}}$ is given by

$$f^{\epsilon,\tilde{H}}((\pi_t)_{t\in[0,T]}) := \langle \pi_T, \tilde{H}_T\rangle - \langle \pi_0, \tilde{H}_0\rangle - \int_0^T \langle \pi_t, \partial_t\tilde{H}_t\rangle\,\mathrm{d}t - \int_0^T\int_\Lambda \phi\big([\pi_t * \iota_\epsilon](u)\big)\Delta\tilde{H}_t\,\mathrm{d}u\,\mathrm{d}t$$
$$+ \int_0^T\int_\Lambda \chi\big([\pi_t * \iota_\epsilon](u)\big)\nabla V\cdot\nabla\tilde{H}_t\,\mathrm{d}u\,\mathrm{d}t - \frac{1}{2}\int_0^T\int_\Lambda \chi\big([\pi_t * \iota_\epsilon](u)\big)|\nabla\tilde{H}_t|^2\,\mathrm{d}u\,\mathrm{d}t.$$

Finally, since the hydrodynamic path $(\pi_t)_{t\in[0,T]}$ solves $\dot\rho_t = \Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla(V+\tilde H_t))$, we obtain

$$f^{0,\tilde H}((\pi_t)_{t\in[0,T]}) = \frac{1}{2}\int_0^T \|\tilde H_t\|^2_{1,\chi(\rho_t)} = \frac{1}{2}\int_0^T \|\dot\rho_t - \Delta\phi(\rho_t) - \nabla\cdot(\chi(\rho_t)\nabla V)\|^2_{-1,\chi(\rho_t)}.$$

□

Acknowledgements We are grateful for stimulating discussions with Federico Cornalba, Max Fathi and André Schlichting. Further, we would like to thank Mark A. Peletier for valuable suggestions. MK is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/L015684/1. JZ gratefully acknowledges funding by the EPSRC through project EP/K027743/1, the Leverhulme Trust (RPG-2013-261) and a Royal Society Wolfson Research Merit Award.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] Giada Basile, Dario Benedetto, and Lorenzo Bertini. A gradient flow approach to linear Boltzmann equations, 2017.

[3] O. Benois, C. Kipnis, and C. Landim. Large deviations from the hydrodynamical limit of mean zero asymmetric zero range processes. *Stochastic Process. Appl.*, 55(1):65–89, 1995.

[4] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Stochastic interacting particle systems out of equilibrium. *J. Stat. Mech. Theory Exp.*, (7):P07014, 35, 2007.

[5] Lorenzo Bertini, Alberto De Sole, Davide Gabrielli, Giovanni Jona-Lasinio, and Claudio Landim. Macroscopic fluctuation theory. *Rev. Modern Phys.*, 87(2):593–636, 2015.

[6] Lorenzo Bertini, Claudio Landim, and Mustapha Mourragui. Dynamical large deviations for the boundary driven weakly asymmetric exclusion process. *Ann. Probab.*, 37(6):2357–2403, 2009.

[7] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[8] Giovanni A Bonaschi and Mark A Peletier. Quadratic and rate-independent limits for a large-deviations functional. *Continuum Mechanics and Thermodynamics*, 28(4):1191–1219, 2016.

[9] David Chandler. *Introduction to modern statistical mechanics*. The Clarendon Press, Oxford University Press, New York, 1987.

[10] Donald A. Dawson and Jürgen Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[11] Nicolas Dirr, Marios Stamatakis, and Johannes Zimmer. Entropic and gradient flow formulations for nonlinear diffusion. *J. Math. Phys.*, 57(8):081505, 13, 2016.

[12] Jean Dolbeault, Bruno Nazaret, and Giuseppe Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations*, 34(2):193–231, 2009.

[13] Manh Hong Duong, Vaios Laschos, and Michiel Renger. Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM Control Optim. Calc. Var.*, 19(4):1166–1188, 2013.

[14] Manh Hong Duong, Agnes Lamacz, Mark A Peletier, and Upanshu Sharma. Variational approach to coarse-graining of generalized gradient flows. *Calculus of Variations and Partial Differential Equations*, 56(4):100, 2017.

[15] Manh Hong Duong, Agnes Lamacz, Mark A Peletier, André Schlichting, and Upanshu Sharma. Quantification of coarse-graining error in Langevin and overdamped Langevin dynamics. *arXiv preprint arXiv:1712.09920*, 2017.

[16] Matthias Erbar, Max Fathi, Vaios Laschos, and André Schlichting. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.*, 36(12):6799–6833, 2016.

[17] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[18] Max Fathi and Marielle Simon. *The Gradient Flow Approach to Hydrodynamic Limits for the Simple Exclusion Process*, pages 167–184. Springer International Publishing, Cham, 2016.

[19] Jin Feng and Thomas G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[20] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Comm. Math. Phys.*, 118(1):31–59, 1988.

[21] Marcus Kaiser, Robert L. Jack, and Johannes Zimmer. Canonical structure and orthogonality of forces and currents in irreversible Markov chains. *J. Stat. Phys.*, 170(6):1019–1050, 2018.

[22] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[23] Elena Kosygina. The behavior of the specific entropy in the hydrodynamic scaling limit. *Ann. Probab.*, 29(3):1086–1110, 2001.

[24] Matthias Liero, Alexander Mielke, Mark A Peletier, and D. R. Michiel Renger. On microscopic origins of generalized gradient structures. *Discrete Contin. Dyn. Syst. Ser. S*, 10(1):1–35, 2017.

[25] Stefano Lisini. *Absolutely continuous curves in Wasserstein spaces with applications to continuity equation and nonlinear diffusion equations*. PhD thesis, Universitá degli Studi di Pavia, 2006.

[26] Stefano Lisini. Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces. *ESAIM Control Optim. Calc. Var.*, 15(3):712–740, 2009.

[27] Stefano Lisini and Antonio Marigonda. On a class of modified Wasserstein distances induced by concave mobility functions defined on bounded intervals. *Manuscripta Math.*, 133(1-2):197–224, 2010.

[28] Jan Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.

[29] C. Maes and K. Netočný. Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states. *Europhys. Lett. EPL*, 82(3):Art. 30003, 6, 2008.

[30] C. Maes, K. Netočný, and B. Wynants. On and beyond entropy production: the case of Markov jump processes. *Markov Process. Related Fields*, 14(3):445–464, 2008.

[31] A. Mielke, M. A. Peletier, and D. R. M. Renger. On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. *Potential Anal.*, 41(4):1293–1327, 2014.

[32] Alexander Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.

[33] Alexander Mielke. On evolutionary Γ-convergence for gradient systems. In *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, volume 3 of *Lect. Notes Appl. Math. Mech.*, pages 187–249. Springer, [Cham], 2016.

[34] L. Onsager and S. Machlup. Fluctuations and irreversible processes. *Physical Rev. (2)*, 91:1505–1512, 1953.

[35] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis.

[36] Filippo Santambrogio. *Optimal transport for applied mathematicians.* Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[37] Sylvia Serfaty. Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *Discrete Contin. Dyn. Syst.*, 31(4):1427–1451, 2011.

[38] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[39] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 2009. Old and new.

4.2. Conclusions

We discussed the convergence of the Ψ-$\Psi^\star$ structure for a class of interacting particle systems to the corresponding quadratic structure on the hydrodynamic scale.

In Section 2, we defined the class of interacting particle systems with gradient dynamics, which we consider in this article. Further, we defined the microscopic and macroscopic quantities associated to these particle systems and derived a representation in terms of the Ψ-$\Psi^\star$ formulas for Markov chains.

In Section 3.1, we derived a chain rule for the microscopic free energy (Proposition 3.1) and stated sufficient conditions for the validity of the chain rule for the macroscopic free energy (Theorem 3.2). In Section 3.2, we stated sufficient conditions for the replacement lemma to hold, which ensure local equilibration of the system.

Section 3.3 contains our main results, which consist of three theorems. Under the assumption that P_L concentrates asymptotically on some $(\pi_t)_{t\in[0,T]}$, we showed in Theorem 3.4 that the macroscopic action functional is controlled by the sequence of microscopic functionals, i.e.

$$\liminf_{L\to\infty} \frac{1}{L^d}\mathcal{H}(P_L|P_L^V) \geq \mathbb{A}\big((\pi_t)_{t\in[0,T]}\big)$$

and similar for the individual terms involved in the Ψ-$\Psi^\star$ representation.

For the special choice $P_L = P_L^{V+\tilde{H}}$ that coincides with P_L^V up to a replacement of the external potential V by a time dependent potential $V + \tilde{H}_t$, this result can be strengthened (see Theorem 3.5) to

$$\lim_{L\to\infty} \frac{1}{L^d}\mathcal{H}(P_L^{V+\tilde{H}}|P_L^V) = \frac{1}{4}\int_0^T \big\|\dot\rho_t - \Delta\phi(\rho_t) - \nabla\cdot(\chi(\rho_t)\nabla V)\big\|^2_{-1,\chi(\rho_t)}\mathrm{d}t.$$

Under the assumption that the chain rule for the macroscopic free energy is satisfied, Theorem 3.6 shows that the structure of Markov chains converges in the limit to the structure of MFT. More precisely, for all $t \in [0,T]$

$$\lim_{L\to\infty} \frac{1}{L^d}\mathcal{F}^V_{L,\alpha}(\mu_t^L) = \mathcal{F}^V_\alpha(\rho_t),$$

as well as

$$\lim_{L\to\infty} \frac{1}{L^d}\int_0^T \Psi_L(\mu_t^L, \jmath_t^L)\,\mathrm{d}t = \frac{1}{2}\int_0^T \big\|\dot\rho_t\big\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t$$

and

$$\lim_{L\to\infty} \frac{1}{L^d}\int_0^T \Psi_L^\star\big(\mu_t^L, F^V(\mu_t^L)\big)\mathrm{d}t = \frac{1}{2}\int_0^T \big\|\Delta\phi(\rho_t) + \nabla\cdot(\chi(\rho_t)\nabla V)\big\|^2_{-1,\chi(\rho_t)}\,\mathrm{d}t.$$

Section 4 contains regularity results for the macroscopic paths and the proof of Theorem 3.2, which yields sufficient conditions for the chain rule. Section 5 contains the remaining proofs of the results in Section 3.3.

Chapter 5

Final Conclusions and Outlook

In this thesis, we considered various result for stochastic systems with a special emphasis on interacting particle systems and their hydrodynamic scaling limits.

In Chapter 2, we discussed the question of acceleration of convergence to equilibrium by breaking time-reversal symmetry. We reviewed well-known acceleration results for ergodic Markov processes (such as Markov chains and single particle diffusions), which show that the relaxation of a given time-reversal symmetric process to its steady state can be accelerated by replacing the dynamics with suitably modified irreversible dynamics. In order to construct these irreversible dynamics, one has to ensure that the new process is ergodic w.r.t. the same distribution. This question has recently been considered in more generality in [10]. We further showed that the above results can be extended to the regime of hydrodynamic scaling limits described by MFT [2].

In Chapter 3, we considered a variational structure for Markov chains [39, 40] and discussed a splitting of the thermodynamic force ($F = F^S + F^A$), for which we proved a prior unknown 'generalised orthogonality' relation. The anti-symmetric part of the force F^A allows to characterise the irreversibility of the process (such that time-reversal symmetric processes satisfy $F^A = 0$). In particular, one can restate 'breaking detailed balance' in Chapter 2 as replacing $F^A = 0$ with a suitable $F^A \neq 0$. We further discussed that this general Ψ-$\Psi^\star$ formula also accounts for MFT, which allows to recover many results available in MFT (such as the *Hamilton-Jacobi orthogonality*, cf. Equation (2.22) in [3]) from properties of the general Ψ-$\Psi^\star$ structure.

In Chapter 4 we demonstrated another connection between the canonical structure for Markov chains and the corresponding structure in MFT. We considered dynamical properties for a family of time-reversal symmetric interacting particle systems of gradient type: We derived a representation of the action functional (the relative entropy on path space) in terms of the free energy and the Ψ-$\Psi^\star$ structure of the Markov chain. Under the assumption of the existence of a hydrodynamic limit, we showed how the quantities for Markov chains converge to their counterparts on the hydrodynamic scale.

5.1. Outlook

Concerning the results in Chapter 2, it would be interesting to see if and how one can design optimal (or practical) irreversible dynamics suitable for sampling purposes, as e.g. considered in [35, 37, 38] for diffusions and jump processes.

In Chapter 4 we have proven results for hydrodynamic limits of time-reversal symmetric particle systems. A natural step would be to try to extend these results to irreversible systems, where $F^A \neq 0$ (which could as well be time-dependent). We also assumed that the limit Q^* concentrates (i.e. $Q^* = \delta_{(\rho_t)_{t\in[0,T]}}$). An interesting extension of our result would be to prove that Q^* concentrates (rather than having this assumption), which might can be done similar to the approaches in [14] or [16]. Without this assumption of concentration, this could lead to an alternative proof for hydrodynamic limits of interacting particle systems. Moreover, it would be interesting to check if our results can be used to extend the results in [14] to irreversible systems.

Bibliography

[1] Lorenzo Bertini, Alberto De Sole, Davide Gabrielli, Gianni Jona-Lasinio, and Claudio Landim. Current fluctuations in stochastic lattice gases. *Physical review letters*, 94(3):030601, 2005.

[2] Lorenzo Bertini, Alberto De Sole, Davide Gabrielli, Giovanni Jona-Lasinio, and Claudio Landim. Macroscopic fluctuation theory. *Reviews of Modern Physics*, 87(2):593, 2015.

[3] Lorenzo Bertini, Alessandra Faggionato, Davide Gabrielli, et al. Large deviations of the empirical flow for continuous time markov chains. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 51, pages 867–900. Institut Henri Poincaré, 2015.

[4] Joris Bierkens. Non-reversible metropolis-hastings. *Statistics and Computing*, 26(6):1213–1228, 2016.

[5] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[6] Anton Bovier. *Statistical mechanics of disordered systems: a mathematical perspective*, volume 18. Cambridge University Press, 2006.

[7] Michael E Cates and Julien Tailleur. Motility-induced phase separation. *Annu. Rev. Condens. Matter Phys.*, 6(1):219–244, 2015.

[8] David Chandler. Introduction to modern statistical mechanics. *Introduction to Modern Statistical Mechanics, by David Chandler, pp. 288. Foreword by David Chandler. Oxford University Press, Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771*, page 288, 1987.

[9] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721, 1999.

[10] Leonardo De Carlo and Davide Gabrielli. Gibbsian stationary non-equilibrium states. *Journal of Statistical Physics*, 168(6):1191–1222, 2017.

[11] Amir Dembo and Ofer Zeitouni. Large deviations techniques and applications, volume 38 of stochastic modelling and applied probability, 2010.

[12] Frank Den Hollander. *Large deviations*, volume 14. American Mathematical Soc., 2008.

[13] Richard S Ellis. *Entropy, large deviations, and statistical mechanics*. Springer, 2007.

[14] Matthias Erbar, Max Fathi, Vaios Laschos, and André Schlichting. Gradient flow structure for mckean-vlasov equations on discrete spaces. *arXiv preprint arXiv:1601.08098*, 2016.

[15] Denis J Evans and Debra J Searles. The fluctuation theorem. *Advances in Physics*, 51(7):1529–1585, 2002.

[16] Max Fathi and Marielle Simon. The gradient flow approach to hydrodynamic limits for the simple exclusion process. In *From Particle Systems to Partial Differential Equations III*, pages 167–184. Springer, 2016.

[17] Étienne Fodor, Cesare Nardini, Michael E Cates, Julien Tailleur, Paolo Visco, and Frédéric van Wijland. How far from equilibrium is active matter? *Physical review letters*, 117(3):038103, 2016.

[18] Mark Iosifovich Freidlin and Alexander D Wentzell. Random perturbations. In *Random Perturbations of Dynamical Systems*, pages 15–43. Springer, 1998.

[19] Giovanni Gallavotti and Ezechiel Godert David Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Physical Review Letters*, 74(14):2694, 1995.

[20] Miroslav Grmela and Hans Christian Öttinger. Dynamics and thermodynamics of complex fluids. i. development of a general formalism. *Physical Review E*, 56(6):6620, 1997.

[21] Thai M Hoang, Rui Pan, Jonghoon Ahn, Jaehoon Bang, HT Quan, and Tongcang Li. Experimental test of the differential fluctuation theorem and a generalized jarzynski equality for arbitrary initial states. *Physical Review Letters*, 120(8):080602, 2018.

[22] Chii-Ruey Hwang, Shu-Yin Hwang-Ma, Shuenn-Jyi Sheu, et al. Accelerating diffusions. *The Annals of Applied Probability*, 15(2):1433–1444, 2005.

[23] Robert L Jack, Marcus Kaiser, and Johannes Zimmer. Symmetries and geometrical properties of dynamical fluctuations in molecular dynamics. *Entropy*, 19(10):562, 2017.

[24] Robert L Jack and Johannes Zimmer. Geometrical interpretation of fluctuating hydrodynamics in diffusive systems. *Journal of Physics A: Mathematical and Theoretical*, 47(48):485001, 2014.

[25] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[26] Christopher Jarzynski. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annu. Rev. Condens. Matter Phys.*, 2(1):329–351, 2011.

[27] Giovanni Jona-Lasinio. Understanding non-equilibrium: a challenge for the future. *Contributions to science*, pages 127–130, 2016.

[28] Marcus Kaiser, Robert L Jack, and Johannes Zimmer. Acceleration of convergence to equilibrium in markov chains by breaking detailed balance. *Journal of Statistical Physics*, 168(2):259–287, 2017.

[29] Marcus Kaiser, Robert L. Jack, and Johannes Zimmer. Canonical structure and orthogonality of forces and currents in irreversible markov chains. *Journal of Statistical Physics*, 170(6):1019–1050, 2018.

[30] Marcus Kaiser, Robert L. Jack, and Johannes Zimmer. A variational structure for interacting particle systems and their hydrodynamic scaling limits. *arXiv preprint arXiv:1805.01411*, 2018.

[31] George Kesidis and Jean Walrand. Relative entropy between markov transition rate matrices. *IEEE transactions on information theory*, 39(3):1056–1057, 1993.

[32] Claude Kipnis and Claudio Landim. Scaling limits of interacting particle systems. 1999.

[33] Anatoly B Kolomeisky and Michael E Fisher. Molecular motors: a theorist's perspective. *Annu. Rev. Phys. Chem.*, 58:675–695, 2007.

[34] Mourtaza Kourbane-Houssene, Clément Erignoux, Thierry Bodineau, and Julien Tailleur. Exact hydrodynamic description of active lattice gases. *arXiv preprint arXiv:1801.08952*, 2018.

[35] Tony Lelièvre, Francis Nier, and Grigorios A Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.

[36] Yu Levin. Internal thermal noise in the ligo test masses: A direct approach. *Physical Review D*, 57(2):659, 1998.

[37] Yi-An Ma, Tianqi Chen, and Emily Fox. A complete recipe for stochastic gradient mcmc. In *Advances in Neural Information Processing Systems*, pages 2917–2925, 2015.

[38] Yi-An Ma, Emily B Fox, Tianqi Chen, and Lei Wu. Irreversible samplers from jump and continuous markov processes. *arXiv preprint arXiv:1608.05973*, 2016.

[39] Christian Maes and Karel Netočnỳ. Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states. *EPL (Europhysics Letters)*, 82(3):30003, 2008.

[40] Christian Maes, Karel Netocny, and Bram Wynants. On and beyond entropy production: the case of markov jump processes. *Markov Processes and Related Fields*, 14(3):445–464, 2008.

[41] Alexander Mielke, Mark A Peletier, and DR Michiel Renger. On the relation between gradient flows and the large-deviation principle, with applications to markov chains and diffusion. *Potential Analysis*, 41(4):1293–1327, 2014.

[42] Alexander Mielke, DR Michiel Renger, and Mark A Peletier. A generalization of onsager's reciprocity relations to gradient flows with nonlinear mobility. *Journal of Non-Equilibrium Thermodynamics*, 41(2):141–149, 2016.

[43] Lars Onsager and S Machlup. Fluctuations and irreversible processes. *Physical Review*, 91(6):1505, 1953.

[44] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.

[45] Francesco Ragone, Jeroen Wouters, and Freddy Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, page 201712645, 2017.

[46] Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081, 2015.

[47] Luc Rey-Bellet and Konstantinos Spiliopoulos. Improving the convergence of reversible samplers. *Journal of Statistical Physics*, 164(3):472–494, 2016.

[48] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012.

[49] Udo Seifert. Stochastic thermodynamics: From principles to the cost of precision. *Physica A: Statistical Mechanics and its Applications*, 2017.

[50] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.